



HARVARD | BUSINESS | SCHOOL

ANYWHERE SIKOCHI
ASSISTANT PROFESSOR OF BUSINESS ADMINISTRATION
BEROL CORPORATION FELLOW

June 8, 2021

Chairman Gary Gensler
U.S. Securities Exchange Commission
100 F Street NE
Washington, DC 20549

Re: Request for Comment on Climate Disclosure

Dear Chairman Gensler:

Thank you for the opportunity to provide input regarding the Securities and Exchange Commission's (SEC) regulation of climate change disclosure. Attached to this letter is a recent a paper I wrote with George Serafeim (Harvard Business School) and Dane Christensen (University of Oregon) titled "Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings." (The final published version of the paper is available at <https://doi.org/10.2308/TAR-2019-0506>).

In this paper we find evidence that increased environmental, social, and governance (ESG) disclosures by companies appears to exacerbate disagreement across rating agencies regarding what sustainability rating to give to individual firms.

If you have any further questions, feel free to reach out.

Sincerely,

Anywhere Sikochi
Harvard Business School

Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings

Dane Christensen
University of Oregon

George Serafeim
Harvard University

Anywhere Sikochi*
Harvard University

December 2020

Abstract

Despite the rising use of environmental, social, and governance (ESG) ratings, there is substantial disagreement across rating agencies regarding what rating to give to individual firms. As what drives this disagreement is unclear, we examine whether a firm's ESG disclosure helps explain some of this disagreement. We predict and find that greater ESG disclosure actually leads to *greater* ESG rating disagreement. These findings hold using firm fixed effects, and using a difference-in-differences design with mandatory ESG disclosure shocks. We also find that raters disagree more about ESG outcome metrics than input metrics (policies), and that disclosure appears to amplify disagreement more for outcomes. Lastly, we examine consequences of ESG disagreement and find that greater ESG disagreement is associated with higher return volatility, larger absolute price movements, and a lower likelihood of issuing external financing. Overall, our findings highlight that ESG disclosure generally exacerbates ESG rating disagreement rather than resolving it.

Keywords: ESG Ratings; Rating Agency Disagreement; ESG Disclosure; Corporate Social Responsibility; Sustainability

JEL Classification: G24, M14, M41, Q56

The data used in this study are publicly available from the sources cited in the text

We thank Vishal Baloria, Mark Bradshaw, Mary Ellen Carter, Jeff Doyle, Jacquelyn Gillette, Michelle Hanlon, Amy Hutton, Zach Kaplan, Kevin Koharki, Brad Lindsey, Christopher Noe, Kyle Peterson, Sugata Roychowdhury, Susan Shu, Chad Simon, Chris Skousen, Eric So, Rodrigo Verdi, Joe Weber, and workshop participants at Boston College, Harvard Business School, MIT, Utah State University, and Washington University in St. Louis and the Bar-Ilan University New Research on Executive Compensation and on Sustainability conference (2019) for useful comments and suggestions. We also thank Carolyn Liu and James Zeitler for research assistance. Christensen is grateful for financial support from the Lundquist College of Business. Serafeim and Sikochi are grateful for financial support from the Division of Research and Faculty Development at Harvard Business School. All errors are our own.

* Corresponding author: Anywhere (Siko) Sikochi, Harvard Business School, Soldiers Field Rd, 389 Morgan Hall, Boston, MA 02163; Telephone: (617) 496-3756; Email: ssikochi@hbs.edu.

I. INTRODUCTION

One of the biggest developments in financial markets, in recent years, has been the integration of environmental, social and governance (ESG) information in investment decisions. According to recent estimates, more than \$30 trillion in assets under management are invested using sustainable strategies that apply ESG criteria in investment analysis and portfolio selection (GSIA 2018). Seeking to capture ESG information that thousands of publicly listed firms increasingly disclose, investor spending on ESG ratings from data providers (i.e., ESG rating agencies) increased from \$200 million to \$500 million between 2014 and 2018 (Gilbert 2019).

In line with the increased use of ESG ratings by market participants, a growing number of academic studies have examined the association of ESG ratings with variables of interest, including stock market performance, accounting performance, financial constraints, and governance characteristics (e.g., Cheng, Ioannou, and Serafeim 2014; Khan, Serafeim, and Yoon 2016; Hubbard, Christensen, and Graffin 2017).¹ However, recently academic research (Chatterji et al. 2016; Berg, Koelbel, and Rigobon 2020) and commentators in the media have questioned these ratings noting the high degree of disagreement among data providers. For example, one *Wall Street Journal* article noted, “Environmental, social and governance criteria are hard to define. When we measure how different ESG providers rate companies in the S&P 500, there's often little overlap. By contrast, when ratings agencies score those same companies for their creditworthiness, they are much more often in agreement” (Sindreu and Kent 2018).² Similarly, regulators have also voiced concerns about the substantial disagreement in ESG ratings. Commissioner Peirce of the Securities and Exchange Commission (SEC) mentioned in a recent speech that “the different

¹ ESG ratings are evaluations of a company based on a comparative assessment of their quality, standard or performance on environmental, social or governance (ESG) issues (Brackley, Petroy, and Wong 2019).

² For a simple illustration of this, see the graphic in Appendix C.

[ESG] ratings available can vary so widely, and provide such bizarre results that it is difficult to see how they can effectively guide investment decisions.”³ These examples highlight the lack of consistency in assigning ESG ratings to companies across different providers.

This issue is of great importance because in the absence of agreement on what good ESG performance constitutes, market participants might be misled by ESG ratings. As an article in the *Financial Times* noted “Investors need to be clear about what the methodology they choose is actually measuring, and why. Otherwise ESG scoring risks creating a false sense of confidence among investors who don’t really understand what lies behind the numbers — and therefore don’t really understand what they’re buying” (Allen 2018).

While several articles highlight the fact that data providers greatly disagree on how to rate a company’s ESG performance, we have very little evidence on why providers disagree so much. Without understanding the reasons for this disagreement, it is difficult to understand not only what the potential remedies could be, but also the plausible consequences of this disagreement.⁴

To address this, we focus on a key firm attribute that we hypothesize is likely to be of first-order importance in causing providers to disagree on ESG ratings. Specifically, we focus on the extent of a firm’s ESG disclosure, as theory suggests that disagreement arises due to different information sets and/or different interpretations of information (e.g., Cookson and Niessner 2020). Conventional wisdom and a plethora of evidence in other settings, such as credit ratings and analyst forecasts, suggest that higher disclosure would be associated with lower disagreement (e.g., Lang

³ <https://www.sec.gov/news/speech/speech-peirce-061819>

⁴ In a concurrent working paper, Kimbrough, Wang, and Wei (2019) find that a measure of ESG disagreement across two rating agencies is positively associated with higher annual return volatility, bid-ask spreads, and analyst forecast dispersion and negatively associated with the presence of a GRI report. This paper uses a substantially smaller sample making it difficult to compare the consistency of our findings. Moreover, although it is plausible that the presence of a GRI report may be associated with the quantity of disclosure, it is problematic as a proxy for the extent of ESG disclosure both because firms disclose significant amounts of ESG information outside GRI reports (e.g., via websites and sustainability or integrated reports that do not follow GRI guidelines) and because the amount of information included in GRI reports varies significantly.

and Lundholm 1996; Morgan 2002; Hope 2003), as greater disclosure reduces information asymmetry. We argue that in our setting, due to the subjective nature of ESG information, the opposite would be true, where higher disclosure would be associated with higher disagreement, as disclosure expands opportunities for different interpretations of information.

Specifically, in the absence of disclosure, data providers are more likely to agree because they use similar rules of thumb and imputation techniques. For example, they likely perceive the lack of disclosure on issues that are widespread and significant for a given industry (Khan, Serafeim, and Yoon 2016) as a bad attribute and thereby assign bad performance to the company. Similarly, for less significant issues where few companies in an industry disclose information, they likely perceive the lack of disclosure as a sign that the issue is relatively unimportant and therefore impute the company's performance to be the average performance of companies in the industry (Kotsantonis and Serafeim 2019).⁵

In contrast, for companies with higher levels of disclosure, data providers need to make a judgment about whether the disclosure means good or bad performance. For example, a company that discloses lost-time injury rates needs to be judged based on this disclosure. This gives rise to a level of subjectivity that leads to higher levels of disagreement. We argue that this subjectivity increases as firms expand their disclosures. This prediction is consistent with arguments from the sociology literature, which theorizes that a plurality of evaluations is likely to occur in newly emerging fields where rules and norms for evaluation are less developed (Lamont 2012). Higher disclosure also increases the likelihood that the ESG rating agencies might be able to use different metrics to evaluate a firm's performance on the same issue and therefore lead to greater rating

⁵ How significant or material an issue is for a given industry is predefined according to the proprietary analysis framework of each rating agency. Therefore, disclosure patterns and associated imputations happen given the classification of whether an issue-industry pair is significant or not.

disagreement.⁶ Moreover, as in financial markets, evaluators may disagree over which measures are more relevant to assessing ESG performance.⁷ Collectively, these arguments suggest that greater ESG disclosure would result in greater ESG rating disagreement.

To test this prediction, we analyze data from three of the largest providers of ESG ratings to investors: MSCI, Thomson Reuters, and Sustainalytics. These rating agencies act as the main information intermediaries for ESG information in financial markets, similar to how credit rating agencies act as information intermediaries and gatekeepers in financial markets. We also utilize ESG disclosure scores from Bloomberg to proxy for the extent of firms' ESG disclosure practices.

Using 30,700 firm-year observations across 5,637 unique firms between the years 2004 and 2016, we find strong support for our prediction that greater ESG disclosure leads to greater ESG rating disagreement. First, we provide descriptive evidence that although ESG disclosure has increased dramatically in the last two decades (through voluntary and mandatory disclosure efforts), the level of ESG disagreement for a given firm has in fact increased over the same period. We corroborate these results using multivariate models that control for a number of other firm characteristics including firm size, profitability, analyst coverage, institutional ownership, industry membership, and valuation multiples. Specifically, we estimate panel regressions with industry and year fixed effects and find a strong positive relationship between ESG disclosure and disagreement. We also estimate models with firm fixed effects and find similarly strong results of a positive association. Overall, these findings suggest that greater ESG disclosure appears to lead to greater ESG rating disagreement.

⁶ Agencies may use different metrics because of different ideologies or, as noted in prior literature (Merton 1987; Hirshleifer and Teoh 2003; Bonsall and Miller 2017), rating agencies could simply focus on different factors because processing all the information contained in disclosures may be too costly.

⁷ In financial markets, Bloomfield and Fischer (2011) highlight that disagreement can arise from investors' differing perceptions regarding the relevance of disclosures.

Next, we consider each of the three pillars E, S, and G separately. Although substantial disagreement still exists across each of these areas regarding what constitutes good performance and what metrics to use, the environmental and social pillars have been debated for a shorter period of time than governance. As disclosure should contribute to greater disagreement where there is less of a shared understanding on what constitutes good performance, we expect disclosures on environmental and social issues to be relatively more likely to contribute to disagreement than governance disclosures. Consistent with this intuition, we find that the environmental and social pillars of ESG disclosure, rather than governance, primarily drive our main results on the relationship between ESG disclosure and disagreement. We observe similar results when we examine the association between pillar-specific disclosure (e.g., social disclosure) and pillar-specific disagreement (e.g., social disagreement).

We also perform supplemental analyses to corroborate our primary inferences. To address potential identification concerns, we use the staggered adoption of broad mandatory ESG disclosure requirements across countries as shocks to firm's ESG disclosures. Using a difference-in-differences design, we find that after a country or stock exchange implements mandatory ESG disclosure requirements, the affected firms tend to experience greater ESG rating disagreement (relative to the control firms). These findings provide greater confidence in our primary inferences.

Next, to better understand what ESG ratings capture and how disclosure contributes to disagreement, we examine individual metrics that ESG raters use to construct their ratings. We categorize the metrics into inputs and outcomes. Inputs refer to efforts that a company is making to achieve a desired outcome (e.g., the presence of a diversity policy) and outcomes refer to actual performance outcomes (e.g., percentage of women in the workforce). We predict and find that ESG raters tend to disagree less about ESG inputs and more about ESG outcomes. This is

consistent with the notion that the evaluation of outcomes is more subjective and relies on having a shared understanding of what a good versus a bad outcome might be. Further, we predict and find that greater ESG disclosure exacerbates these disagreements, especially in the case of outcomes, consistent with our argument that more pieces of information that require subjective evaluation should lead to greater disagreement. Moreover, we find that greater disclosure of ESG information at the industry-level also exacerbates ESG rating disagreement when evaluating ESG outcomes, suggesting that widespread increases in ESG disclosure across firms is unlikely to resolve ESG disagreement until there are common norms for what constitutes good and bad ESG performance.

Lastly, we explore market consequences of ESG rating disagreement. Using three-day window tests around the release of revised ESG ratings ($t-1$, $t+1$), we find greater ESG disagreement is associated with higher stock return volatility and larger absolute price movements. These findings hold both with and without firm fixed effects, and suggest that ESG disagreement is relevant to market participants and influences stock prices. We also find some evidence that these results are becoming even stronger over time, which suggests that ESG disagreement is having an increasing impact on markets. In addition, we explore the influence of ESG disagreement on firms' financing choices. Consistent with ESG disagreement creating market frictions by introducing uncertainty regarding a firm's long-term sustainability, we find that firms with greater ESG disagreement are less likely to raise external financing and instead tend to rely more on internal financing.

Overall, our study makes contributions to two streams of literature. First, we contribute to the literature that documents the presence of significant disagreement among ESG data providers (Chatterji et al. 2016). Our results indicate that this disagreement is most pronounced for firms

with high levels of ESG disclosure, thereby shedding light on a key driver of ESG rating disagreement. More broadly, our results contribute to the growing literature that uses ESG data to understand their relationship with other important organizational and market outcomes (Khan, Serafeim, and Yoon 2016; Ferrell, Liang, and Renneboog 2016; Bereskin et al. 2018).

Second, we contribute to the literature that investigates rating disagreement in other settings, such as in credit ratings or analyst forecasts (e.g., Morgan 2002; Lang and Lundholm 1996; Bonsall and Miller 2017; Akins 2018). Contrary to evidence in these settings that disclosure mitigates disagreement, in our setting, disagreement is larger when firms have higher levels of disclosure. In our view, this highlights the importance of developing a shared understanding of a) what constitutes good or bad ESG performance, and b) what metrics to use to capture ESG performance, as preconditions for ESG disclosure to decrease disagreement.

Our findings also illuminate some of the challenges ESG rating agencies face as information intermediaries. While thousands of companies now claim to integrate ESG issues in their business strategy and operations, it is not clear whether those claims are merely ‘cheap talk’ (Delmas and Burbano 2011).⁸ In the presence of information asymmetry and incomplete information about a firm’s ESG performance, ESG ratings can perform a significant information intermediary function, mitigating the adverse selection problem and thereby helping investors and other stakeholders to choose companies that exhibit their preferred ESG outcomes (Chatterji and Toffel 2010). In particular, having rating agencies focus on ESG outcomes might be desirable to mitigate ‘cheap talk’ by companies, as a company would need to show real effects (e.g., reductions in carbon emissions, improvements in lost time injury rates) instead of disclosing the adoption of a policy or initiative that might generate no real effects (Grewal and Serafeim 2020). However,

⁸ For example, many companies have adopted policies about deforestation or diversity. However, deforestation has not slowed down, nor have most companies become more diverse.

our results suggest that ESG rating agencies have a more difficult time agreeing when evaluating a company on outcomes rather than input metrics, and that disclosure exacerbates disagreement on outcomes even more than it does on inputs. This lack of consensus about how to interpret outcomes might be an obstacle that encourages raters to focus more on inputs and thereby causes potential damage to the corporate accountability function that ratings could perform. In this context, we regard disagreement as inhibiting the accountability process. Moreover, while disclosure may have many positive effects, it likely needs to be placed in a framework that allows analysts to evaluate those outcomes with clear benchmarks.⁹

Overall, given concerns over ESG rating disagreement, our findings suggest a lot of work still needs to be done to develop rules and norms to determine what characterizes good ESG performance.

II. INSTITUTIONAL BACKGROUND AND HYPOTHESIS DEVELOPMENT

ESG Rating Agencies

Several ESG data providers have emerged in the last two decades, most of which provide aggregate ratings of a firm's ESG performance. Responding to the need for collecting, interpreting, aggregating, and distributing ESG data, ESG rating agencies have become important information intermediaries in capital markets. The rating agencies share a common objective, which is to measure the ESG performance of a company (MSCI 2018; Thomson Reuters 2017; Sustainalytics 2018). The concept of ESG performance is intended to provide an assessment of how well a company is managing environmental, social, and governance risks and opportunities (MSCI 2018;

⁹ A recent innovation that could serve as a useful example for framework and benchmark setting is the creation and adoption of science-based climate targets from more than 1,000 organizations so far. Organizations need to set their carbon emission reduction targets in line with the Paris Agreement based on the carbon budget remaining to avert an increase in the temperature by 2 degrees Celsius.

Thomson Reuters 2017; Sustainalytics 2018). Although this captures a variety of heterogeneous issues, conceptually it is a summary measure of non-financial performance, which is somewhat analogous to how net income is a summary measure of financial performance covering a variety of heterogeneous areas (e.g., sales, impairments, interest and tax expenses).

The rating agencies assess ESG performance using a wide array of metrics pertaining to a firm's efforts (i.e., policies or programs) or outcomes on ESG issues. Each of the rating providers cover very similar ESG issues, although sometimes they label the issues with different names. The exact metrics that rating agencies cover can differ, and how those metrics are used is a proprietary part of the rating process and is not observable to outside researchers.

Investors are the main audience of the ESG rating agencies featured in our study. The ESG ratings and the underlying data offered by the rating agencies are intended to help investors integrate ESG factors into their decisions, screen portfolios for risks and opportunities, generate investment ideas, conduct due diligence, determine opportunities for engagement, and support implementation of the UN PRI principles (MSCI 2018).¹⁰ ESG rating agencies are compensated by investors, and not by the companies they rate, thereby mitigating potential conflicts of interest.

Of the ESG rating agencies, MSCI is widely considered the largest data provider to the investment community. It sells ESG ratings to investors and uses ESG data to construct stock market indices. Sustainalytics also sells its ESG ratings to investors, and provides other advisory and research services as well. Moreover, following its acquisition by Morningstar, Sustainalytics ratings form the basis for fund-level ESG ratings. Finally, ASSET4 was acquired by Thomson Reuters and its ESG ratings were integrated in Thomson Reuters' platforms and provided to their

¹⁰ In contrast to credit ratings agencies, which often produce alphabetical ratings, historically few ESG rating agencies have translated their scores to alphabetical ratings. Instead they produce numeric cardinal ratings.

subscribers.¹¹ Each of these rating agencies employs over 150 ESG analysts who collect and evaluate ESG data to produce ESG ratings. ESG rating agencies collect relevant information for rating assessments, usually annually, while if there are major events unfolding, they might revise a rating before that. The latter is fairly infrequent. Different companies receive ratings on different months and days of the year. To collect the necessary data they analyze corporate disclosures, such as sustainability reports and corporate websites, administer surveys for companies to complete, and review information coming from other stakeholders, such as regulatory agencies, industry associations and NGOs.¹² However, despite their best efforts, the ESG ratings they produce tend to differ greatly for a given firm across rating agencies, which has drawn substantial criticism from outside observers.

ESG Rating Disagreement

It should be no surprise that ESG ratings would exhibit high levels of disagreement. A large body of work in sociology shows that human ability to make sense of information in a common way occurs over time, and that both cultural and social processes define and enable the evaluation of knowledge (Espeland and Stevens 1998; Knorr-Cetina 1999). This can include determining what information to focus on, how to interpret it, and how much weight to give it.¹³ In areas that are not highly formalized or that are newly emergent, it is more likely that pluralistic evaluative cultures prevail, leading to higher disagreement among raters (Lamont 2012), as in the case of the ESG field. In comparison to financial analysis that has been taught and formalized in the last

¹¹ In October 2018 Thomson Reuters sold a large stake in its data and analytics unit (Refinitiv), which distributes ASSET4 ratings, to Blackstone (Thomson Reuters 2018). Less than a year later, the London Stock Exchange agreed to acquire Refinitiv, although the deal is still awaiting regulatory approval (Isaac and Dummett 2020).

¹² We use timestamped ratings by MSCI and Sustainalytics thereby mitigating any bias arising from ex-post restatements of ratings. In the case of Thomson Reuters because the ratings do not have timestamps it is plausible that ratings are restated based on subsequent information.

¹³ With that said, findings from concurrent research by Berg et al. (2020) suggest that using different weights (i.e., weight divergence) plays a smaller role in ESG rating disagreement than selecting different categories (i.e., scope divergence) and using different metrics (i.e., measurement divergence).

century, ESG analysis has only emerged in the last two decades with most education and formalization occurring in the last ten years.

Moreover, because the information flow of ESG data is less formalized than the flow of financial data, which is systematized by institutional arrangements such as earnings calls and investor presentations, it can lead to further increases in disagreement. As analysts are more likely to receive and evaluate ESG information in a less structured way, analysts might seek and retrieve different pieces of information at different points in time and in a different sequence. This in turn leads to formation of differential expectations about a firm's ESG performance and therefore to different interpretations of subsequent information (e.g., Jones et al. 1968; Krüger and Nolte 2016). In other words, analysts' prior evaluations of specific ESG data can influence their subsequent evaluations of other ESG data, leading to heterogeneity in judgements (Lamont 2009). For example, a prior assessment of a firm's workplace practices might affect an analyst's evaluation of product safety.

In addition, evaluative practices that lead to ratings are influenced by conventions (Becker 1982). Whether evaluators follow customary rules is associated with how strongly they are invested in what defines a proper evaluation and, ultimately, their self-concept as an evaluator. For example, accountants, financial analysts, and credit analysts have established a strong identity through professional associations. No such identity yet exists for ESG analysts. Such a less institutionalized field will be less consistent in providing clear rules and in socializing new evaluators, thereby giving rise to heterogeneity in judgements (Lamont 2012).

ESG Disclosure and Disagreement

With these things in mind, we now turn to the role that ESG disclosure may play in explaining ESG rating disagreement. Theory suggests that disagreement in financial markets

generally arises due to individuals having different information sets and/or different interpretations of information (e.g., Cookson and Niessner 2020). Accordingly, ESG disclosure seems like a natural factor that could influence ESG rating disagreement. Existing literature has shown that disclosure improves a firm's information environment and resolves uncertainties about the firm for participants in both the equity and debt markets.¹⁴ In the equity markets, Lang and Lundholm (1996) provide evidence that a firm's disclosure practices lead to more accurate analyst earnings forecasts, less dispersion among individual analyst forecasts, and less volatility in forecast revisions. Moreover, their evidence suggests that disclosure reduces estimation risk and information asymmetry. Several subsequent studies corroborate these conclusions. For example, Hope (2003) finds that disclosure about a firm's accounting policies reduces uncertainty about its forecasted earnings and thus reduces forecast dispersion and forecast error. Overall, evidence suggests that disclosure reduces disagreement (i.e., forecast dispersion) among equity analysts.

Similarly, in the debt markets, research finds that disclosure appears to reduce disagreement between credit rating analysts' assessment of credit risk of firms. Morgan (2002) finds that credit ratings agencies disagree more often over firms that are more opaque, and Bonsall and Miller (2017) find that firms with more readable disclosures are less likely to have split credit ratings. Additionally, Akins (2018) finds that firms with greater financial reporting quality are less likely to have split credit ratings. This evidence suggests that enhanced transparency through disclosures can mitigate rating disagreement on the premise that split ratings can arise when credit rating agencies have limited access to data with which to base their credit assessments.

We argue that in the context of ESG ratings the opposite would be true, where more

¹⁴ Although research has also found that greater disclosure can temporarily increase disagreement over short windows, such as around earnings announcements (Bamber et al. 1999; 2011), this disagreement is usually resolved shortly thereafter, ultimately leading to lower disagreement (Amiram, Owens, and Rosenbaum 2016). In contrast, in our setting we examine the relation between disclosure and disagreement over a longer window.

disclosure would be associated with higher disagreement. In contrast to financial disclosures where there is widespread agreement about the meaning of specific financial variables, such as leverage or profitability on the future creditworthiness of a company, for ESG disclosures there is no shared understanding yet on which exact metrics should be assessed to evaluate firm's ESG performance, or how to interpret and judge their meaning about a firm's ESG performance.

Anecdotal evidence suggests that this might be true. For example, when Workday Inc. significantly increased its ESG disclosure for fiscal year 2015 by adopting the new G4 Global Reporting Initiative guidelines,¹⁵ its sustainability report increased from 54 to 98 pages. Its ESG disclosure score from Bloomberg increased significantly for both the environmental and social pillars. However, disagreement increased as its Thomson Reuter's rating for both environmental and social issues increased, MSCI's environmental rating decreased and its social rating increased, and Sustainalytics left its environmental rating unchanged while its social rating increased. We observed similar patterns when analyzing other companies that substantially increased their ESG disclosure, such as AT&T, Tech Mahindra and Altria, among others.

The rating disagreements can arise because more disclosure is more likely to lead to rating agencies using different metrics in assessing a firm's ESG performance. For example, on the issue of workplace safety, if a firm only discloses lost time injury rates, then all raters will likely use this metric to assess a company's performance. However, if a firm also discloses additional information (e.g., the number of fatalities due to accidents, the number of lost workdays, time loss claims), raters might use different metrics to assess a firm's safety performance or might assign different importance to the different metrics.¹⁶

¹⁵ The Global Sustainability Standards Board (GSSB) issued the G4 Global Reporting Initiative guidelines. The GSSB is an independent international organization whose core products are the Sustainability Reporting Standards.

¹⁶ For example, this can occur when raters have different ideologies about which measures are relevant to assessing ESG performance (e.g., Bloomfield and Fischer 2011).

In addition, analysts might disagree on how to interpret a given metric. Specifically, in the absence of clear rules for evaluative practices, it is likely that rules of thumb will be developed for the least demanding evaluative tasks. For example, while an ESG analyst would need to judge a specific piece of information, in the absence of that information, a simple rule can be developed. Therefore, in the absence of disclosure, raters are more likely to agree because they use similar rules of thumb and imputation techniques. For instance, as some ESG analysts shared with us, if a single company does not disclose information on an important issue for its industry, they perceive the lack of disclosure as a bad attribute and thereby assign bad performance to the company. Similarly, in the absence of disclosure across most companies in an industry, they likely perceive the lack of disclosure as a sign that the issue is relatively unimportant, thereby imputing the company's performance to be the average performance of companies in the industry.

In contrast, for companies with higher levels of disclosure, ESG analysts need to make a judgement about whether the information being disclosed means good or bad performance and assess how to aggregate the different disclosures that a firm provides. For example, a company that discloses cubic meters of water consumed in operations or lost time injury rates needs to be judged based on these disclosures. This gives rise to a level of subjectivity that leads to ratings disagreement. We argue that this level of subjectivity increases as firms expand their disclosure, since firms first tend to disclose the presence or absence of policies, practices, and strategies (e.g., policies to provide employees with child care support or to require a health and safety policy) and then eventually disclose more specific quantitative metrics (e.g., employee turnover, CO₂ emissions, water recycled). The latter are likely to give rise to the highest disagreement, as they need to be judged based on what constitutes "good" or "bad" performance, while at least in the context of policies, practices, and strategies, more disclosure might mean better performance.

Collectively, all these arguments suggest that greater ESG disclosure increases ESG rating disagreement, because greater disclosure provides more information raters can disagree about, and thus creates more opportunities for raters to have different interpretations of information (Cookson and Niessner 2020). Therefore, we hypothesize:

H1: ESG disclosure is positively associated with ESG rating disagreement.

III. RESEARCH DESIGN AND SAMPLE CONSTRUCTION

Empirical Model

To test our hypothesis that greater ESG disclosure will lead to greater ESG rating disagreement, we estimate the following ordinary least squares (OLS) regression model:

$$ESG_Disagreement_{it} = \beta_0 + \beta_1 ESG_Disclosure_{it} + \sum \beta_k Controls + \varepsilon_{i,t} \quad (1)$$

where *ESG_Disagreement* is the standard deviation of ESG ratings firm *i* received from rating agencies for its performance in year *t*.¹⁷ *ESG_Disclosure* is the firm's ESG disclosure score pertaining to year *t*'s performance. Note that each of these ESG variables is publicly released in year *t*+1, but pertain to year *t*'s performance, with a firm's ESG disclosure occurring before the ESG rating assessment. Thus, the model allows for an assessment of how ESG disclosures influence subsequent ESG disagreement.

Controls consists of the following firm characteristics: the average ESG rating a firm received from various rating agencies for year *t*'s performance (*ESG_Avg*), firm size (*Firm Size*), firm performance (*ROA*), growth opportunities (*BTM*), capital structure (*Leverage*), the number of analysts following the firm (*Analyst Following*), and the percentage of shares held by institutional

¹⁷ We measure ESG disagreement using the standard deviation of ratings, as opposed to using the coefficient of variation, because we measure all the ESG ratings in our sample on the same scale (0 to 100) and scaling by a firm's average rating would complicate inferences and induce a mechanical relationship between a firm's average rating (*ESG_Avg*) and our dependent variable (Sørensen 2002). We also construct an alternative measure of ESG rating disagreement, measured as the average of absolute values of the difference between pairs of ratings that a firm receives for its performance in year *t*. Our inferences are generally similar using this alternative measure (untabulated.)

investors (*Inst. Ownership*).¹⁸ See Appendix A for further details on how each variable is measured. We also include industry, year, country, and rating agency fixed effects. For comparison, in some specifications we replace industry and country fixed effects with firm fixed effects to examine within-firm variation in ESG disagreement.¹⁹ In all specifications, we winsorize all continuous variables at the 1st and 99th percentiles and standard errors are clustered by firm. Our hypothesis predicts that the estimated coefficient on *ESG_Disclosure* will be positively associated with *ESG_Disagreement* (i.e., $\beta_1 > 0$).

Sample Construction

To construct our sample, we obtained ESG ratings from three of the world's most prominent ESG rating agencies, namely Morgan Stanley Capital International's (MSCI) Intangible Value Assessment, Thomson Reuters' (TR) ASSET4, and Sustainalytics.²⁰ These ESG rating agencies provide international coverage, which we later exploit to better get at identification.

As some ESG rating agencies only provide one rating per firm each year, we construct a firm-year dataset by linking the ESG ratings to firm-year observations from Worldscope. When a rating agency release multiple ESG ratings for a given firm-year's performance, we retain the last rating issued within 12 months of the firm's fiscal year-end. This ensures that all the rating agencies in our sample have had an opportunity to observe any ESG disclosures a firm has made relating to year t 's ESG performance.²¹ To ensure that ratings from these agencies are comparable,

¹⁸ Inferences are unchanged if we also include stock return volatility as a control variable.

¹⁹ In the models with firm fixed effects, the reported intercept captures the average of the fixed effects.

²⁰ We do not include MSCI's KLD ratings in our analyses because MSCI discontinued these ratings and they pertain almost exclusively to U.S. firms. However, our inferences are unchanged if we also include these ratings (untabulated). Some studies also use membership in the Dow Jones Sustainability Index, Calvert Index, and FTSE4Good Index as proxies for good ESG performance (e.g., Christensen 2016; Chatterji et al. 2016). These organizations do not publicly release ESG ratings, so we do not view them as rating agencies. Hence we do not include them in our analyses.

²¹ For example, if a firm has a December 31, 2010 year-end (i.e., fiscal year 2010), we use TR ASSET4's 2010 rating, and retain the last ESG rating MSCI and Sustainalytics released between January 1 and December 31, 2011. This approach also helps to mitigate capturing disagreement arising from differences in the information sets of rating agencies. Nevertheless, some differences in the information sets of each ESG rating agency may still exist.

we only use the ratings for the environmental, social, and governance pillars (as some raters also evaluate economic performance).²² We re-scale the ratings when necessary so that they all range from 0 to 100. Specifically, while TR ASSET4 and Sustainalytics scores range from 0 to 100, MSCI's scores range from 0 to 10. Thus, we multiply MSCI's scores by 10 to make them comparable.

To capture the extent of firms' ESG disclosure, we use ESG disclosure scores provided by Bloomberg. These disclosure scores are based on information firms disclose in various ways, such as via sustainability reports, annual reports, and corporate websites.²³ The scores range from 0.1 for firms that disclose a minimum amount of ESG data to 100 for firms that disclose every data point that Bloomberg collects.²⁴ Thus, the more ESG information a company discloses, the higher the disclosure score. Bloomberg's ESG disclosure coverage includes over 10,000 common stocks around the world and as a result is significantly larger (i.e., nearly double) than the coverage universe of common stocks of the ESG rating agencies. Bloomberg also tailors the disclosure scores for different industry sectors, so that companies are only evaluated based on data that is relevant to their industry. We use Bloomberg's overall ESG disclosure score in our primary analyses, similar to prior studies (e.g., Baldini et al. 2018; Li et al. 2018). In supplemental analyses, we also employ the more granular disclosure scores Bloomberg provides for each ESG pillar.

Our sample begins in 2004, the first year that we have data on the extent of firm's ESG

²² Because TR ASSET4's composite score also includes an economic component, we exclude this and calculate the overall ESG score as the average environmental, social, and governance score.

²³ ESG disclosures tend to be less timely than financial disclosures. It is common for firms to disclose sustainability reports and ESG information on their websites several months after they release their annual reports. This likely reflects a combination of the difficulty of collecting some of the ESG information, the smaller human and information system resources dedicated to collecting ESG information compared to collecting financial information, and the lack of regulation mandating the timeliness of this information.

²⁴ Although Bloomberg's ESG disclosure scores are used by practitioners and thus are institutionally relevant, Bloomberg may not capture all relevant ESG disclosures made by firms, so this proxy may measure ESG disclosure imperfectly. Nevertheless, the dynamic difference-in-difference analysis that we perform later in the paper provides some validation for this data.

disclosures, and ends in 2016 based on data availability at the time we constructed our sample. We obtain firm financial data from Worldscope, analyst coverage from the Institutional Brokers' Estimate System (I/B/E/S), and institutional ownership from Factset.

As we are interested in studying ESG rating disagreement, we require each firm-year to have ratings from more than one ESG rating agency. After requiring basic firm-level controls, along with ESG disclosure scores, our final sample consists of 30,700 firm-year observations from 5,637 unique firms. Table 1 outlines a summary of our sample construction process.

IV. DESCRIPTIVE ANALYSES

Descriptive Statistics

Table 2 Panel A provides descriptive statistics illustrating the worldwide coverage of our sample. The sample is comprised of firms from 69 countries, with roughly three-quarters of the observations coming from nine countries: the United States, Japan, the United Kingdom, China, Canada, Australia, France, Germany, and Switzerland.

Table 2 Panel B provides descriptive statistics for the ESG ratings from each of the rating agencies in our sample. While our overall sample has 30,700 firm-year observations, each rating agency has fewer observations than this, as we only require each firm-year to be rated by at least two rating agencies, similar to research on analyst disagreement (e.g., Sadka and Scherbina 2007; Barinov 2013). In our sample, each of the raters tends to issue overall ESG scores of around 50 (out of 100 points possible). With that said, MSCI tends to issue slightly lower overall ESG ratings (48.4), and Sustainalytics tends to issue slightly higher overall ESG ratings (57.2). Additionally, ratings provided by TR tend to have the greatest variance across firms (i.e., standard deviation of 24.5), while ratings provided by Sustainalytics tend to have the least variance across firms (i.e., standard deviation of 9.8). Similar trends exist for each of the underlying environmental, social,

and governance pillars.

Table 2 Panel C provides descriptive statistics for ESG rating disagreement that exists for a given firm-year. While the average overall ESG rating a firm receives for a given year's performance is 52.8 (out of 100), there appears to be substantial disagreement across raters regarding what rating to give that firm. Specifically, the standard deviation of overall ESG ratings that a firm receives for a given firm-year is 12.3 (i.e., *ESG_Disagreement*). To provide a more intuitive sense for the magnitude of this disagreement, we also document the absolute difference in ratings that pairs of agencies provide for a given firm-year. For example, we see that the ratings provided by TR and MSCI for a given firm-year differ by 19.7 points, on average. Similarly, the ratings TR and Sustainalytics provide for a given firm year differ by 15.8, on average. While the ratings from MSCI and Sustainalytics tend to differ by the least amount for a given firm-year (12.7), that difference is also quite large. Fairly similar patterns exist for each of the underlying environmental, social, and governance pillars as well. Overall, this descriptive evidence suggests that there is considerable disagreement across rating agencies regarding how to view a given firm's ESG performance.²⁵

Table 2 Panel D provides descriptive statistics for the variables used in our regression analyses. Here we see that the average firm's ESG disclosure score is 28.5 (out of 100), with a reasonable amount of variation across firms (i.e., a standard deviation of 15.0) consistent with the literature (e.g., Baldini et al. 2018).²⁶ We also see that the average firm in our sample is fairly large

²⁵ In untabulated analyses, we also examine *ESG_Disagreement* by industry and find that there tends to be substantial disagreement across all Fama French 49 industries, with the tobacco industry having slightly higher disagreement than the other industries. If we examine *ESG_Disagreement* by country, we also find there tends to be substantial disagreement across almost all countries (untabulated). No particular country stands out as having much higher disagreement than the others. The few countries without much disagreement tend to have less than 10 observations (e.g., Kazakhstan, Ukraine, Kenya).

²⁶ Interestingly, the disagreement for governance ratings appears to be slightly higher than for environmental and social ratings suggesting a significant level of disagreement even for that pillar, which has been debated for a longer period in rating agencies.

(total assets of US\$ 7.3 billion), profitable (ROA of 0.05), and is followed by around 11 sell-side equity analysts.

Univariate correlations (reported in the online appendix) also illustrate that ESG disagreement tends to be greater for larger firms, firms with better ESG disclosure, and firms with greater analyst following. Another interesting observation is that although *ESG_Disclosure* and *ESG_Avg* are positively correlated with each other, they have very different correlations with *ESG_Disagreement*.²⁷ Specifically, *ESG_Disclosure* is positively correlated with *ESG_Disagreement*, while *ESG_Avg* is negatively correlated with *ESG_Disagreement*.

Has ESG rating disagreement declined over time?

While the descriptive evidence presented thus far suggests that rating agencies tend to have strikingly different views on a given firm's ESG performance, it is possible that this disagreement has been declining over time and is not really much of an issue any more. This could occur if rating agencies' views begin to converge over time regarding what constitutes good ESG performance. This is plausible given efforts over time by such organizations as the GRI and SASB to create sustainability standards.²⁸ Thus, before moving onto the tests of our main hypothesis, we perform a few additional descriptive analyses.

First, we construct a time trend measure where we take the fiscal year for a particular firm-year observation and subtract the first year of our sample (*Time*). We then regress *ESG_Disagreement* on *Time* to examine whether ESG disagreement has been declining over time. As raters have been expanding their coverage over time (e.g., picking up smaller firms), we also include firm fixed effects in the model to ensure that changes in the sample composition over time

²⁷ Variance inflation factors for all the variables in our models are below 3, indicating that multicollinearity is not a problem (Kennedy 2008).

²⁸ The Global Reporting Initiative (GRI) and Sustainability Accounting Standards Board (SASB) are independent non-profit organizations that develop sustainability reporting standards.

don't confound our inferences. This effectively allows us to see how ESG disagreement for a given firm has been changing over time.

As reported in Table 3 Panel A we see that the coefficient on *Time* is positive and statistically significant, suggesting that ESG disagreement for a given firm has actually been increasing over time. This finding is inconsistent with rating agencies' views converging over time regarding what constitutes good ESG performance.²⁹ To illustrate how this trend compares to trends in ESG disclosure, as well as average ESG performance ratings, we next replace the dependent variable with *ESG_Disclosure* and *ESG_Avg*, respectively, and re-run this analysis. In Table 3 Panel B we see that ESG disclosures also appear to be getting better over time (i.e., *Time* is again positive and statistically significant). And in Table 3 Panel C we see that average ESG ratings for a given firm have also been increasing over time.

Overall, these analyses suggest that despite improving ESG disclosures and ESG performance over time, ESG disagreement appears to be increasing over time. Next, we explore whether these trends are actually related.

V. EMPIRICAL RESULTS

What drives ESG rating disagreement?

We now turn to examining our primary hypothesis regarding whether ESG disclosure plays a role in influencing ESG ratings disagreement.

Table 4 reports the results of our tests based on Equation 1 using two different specifications: (1) without firm fixed effects, and (2) with firm fixed effects. Consistent with our

²⁹ After our sample period, Thomson Reuters revised its ESG rating methodology. As they released ratings using the new methodology that also pertain to prior years, we re-ran this analysis using these new ratings as an input into the calculation of *ESG_Disagreement*. The results are very similar (untabulated). Results from tests of our main hypothesis are also very similar using these new ratings (untabulated).

hypothesis that ESG disclosure will be positively associated with ESG disagreement, we find that the estimated coefficient on *ESG_Disclosure* is indeed positive and statistically significant across all the models. These findings suggest that when a firm increases its ESG disclosure, this appears to exacerbate ESG disagreement across rating agencies. In other words, greater disclosure appears to give rating agencies simply more information to disagree about and interpret differently. In terms of the magnitude of this effect, we observe that increasing *ESG_Disclosure* from the 25th to the 75th percentile is associated with a 22 to 31 percent increase in *ESG_Disagreement*, depending on the model specification. This appears to be an economically significant effect.

In terms of other factors that help explain ESG disagreement, we find that the estimated coefficient on *ESG_Avg* is negative and statistically significant across all specifications. These findings suggest that ESG rating agencies tend to disagree more about firms with poor average ESG performance. This is similar to findings from the credit ratings literature, as there tends to be greater disagreement between credit raters regarding firms with lower average credit ratings (Cantor and Parker 1994; Bonsall et al. 2017). In terms of economic significance, we observe that increasing *ESG_Avg* from the 25th to the 75th percentile is associated with a 25 to 32 percent decrease in *ESG_Disagreement*, depending on the model specification. Again, this appears to be an economically meaningful effect, relative to the sample average for *ESG_Disagreement*. Most of the other firm characteristics in the model do not yield consistently significant results.³⁰

We also find that firm fixed effects appear to explain a sizable amount of the variation in ESG disagreement. Specifically, moving from the model without firm fixed effects, to the model

³⁰ The exceptions to this are *BTM* and *Institutional Ownership*, which are negative and significant across the specifications. Although these findings suggest that firms with lower growth opportunities and greater institutional ownership tend to have lower ESG disagreement, the economic magnitude of these findings is relatively small. Increasing *BTM* (*Institutional Ownership*) from the 25th to the 75th percentile is associated with a one to two (four to five) percent decrease in ESG disagreement, depending on the model specification.

with firm fixed effects, we see the explanatory power of the model improves dramatically (i.e., the adjusted R^2 goes from 0.118 to 0.516) suggesting that a lot of the disagreement about a firm's ESG performance is very persistent over time.

What dimensions of ESG disclosure contribute to ESG disagreement?

Next, we examine whether some attributes of the ESG disclosures play a greater role than others do in explaining ESG disagreement. Although substantial disagreement still exists across each of three pillars E, S, and G regarding what constitutes good performance and what metrics to use, the environmental and social pillars have been debated for a shorter period of time than governance. Thus, *ex ante* we expect disclosure to have a *relatively* stronger association with environmental and social rating disagreement compared with governance rating. As disclosure should contribute to greater disagreement where there is less of a shared understanding on what constitutes good performance, we expect disclosures on environmental and social issues to be relatively more likely to contribute to disagreement than governance disclosures. We acknowledge, however, that there is still substantial debate about the measurement and meaning of good corporate governance (Daines, Gow, and Larcker 2010; Guest and Nerino 2020).

Using the separate disclosure scores Bloomberg provides for each ESG pillar, we test which ESG pillars contribute to the positive association between ESG disclosure and ESG disagreement.³¹ Specifically, we re-run our main analysis and replace *ESG_Disclosure* with its underlying component scores (*E_Disclosure*, *S_Disclosure*, and *G_Disclosure*). As reported in Table 5, we find that across the specifications, *E_Disclosure* and *S_Disclosure* have a consistently

³¹ Although we would like to perform textual analyses on firms' ESG disclosures (e.g., sustainability reports) to investigate this issue, we are unfortunately unable to obtain the ESG disclosures made by the firms in our sample. While firms post their most recent sustainability reports on their websites, they generally do not post reports that are over a few years old. Thus getting access to historical reports is problematic. Although CorporateRegister.com has gathered a fairly comprehensive collection of sustainability reports over time, paid subscribers are only allowed to download up to 100 sustainability report pdfs a month. Given the large size of our sample, it is not feasible to gather the reports that pertain to our sample.

positive and significant association with *ESG_Disagreement*. In contrast, *G_Disclosure* is only statistically significant in the model without firm fixed effects. These findings are consistent with our expectations and suggest that firms' environmental and social disclosures appear to contribute to ESG rating disagreement, while firms' governance disclosures appear to generate disagreement only in the cross-section but not over time for a given firm.

Next, we corroborate this more directly by seeing whether disclosure for a particular pillar (e.g., environmental disclosure) helps explain disagreement for that pillar (e.g., environmental disagreement), while controlling for the average performance rating for that pillar (e.g., environmental performance). As shown in the other columns of Table 5, the results from these tests produce similar inferences, again suggesting that more environmental and social disclosures tend to contribute to environmental and social rating disagreement, with governance disclosures not playing as much of a role in governance disagreement.

VI. ADDITIONAL ANALYSES

Shocks to ESG Disclosure

One potential concern regarding our results is that they may be influenced by self-selection or firm-level time-varying correlated omitted variables because firms can choose how much ESG information to disclose. Although we try to mitigate the possibility of such biases by estimating models with firm fixed effects, it still could be the case that our results might be influenced by unobservable variables. To mitigate this concern, we exploit the passage of broad mandatory ESG disclosure requirements, which went into effect in numerous countries during our sample period. These broad ESG disclosure requirements arose due to legislation at the country level or listing requirements from individual stock exchanges. From an identification perspective, the attractive features of these mandatory disclosure requirements are that they are staggered across many

countries over time, and they should lead to exogenous increases in ESG disclosure. A summary of the laws/regulations that were passed in different countries is shown in Appendix B.³² Based on these laws/regulations, we create an indicator variable equal to one if the firm-year's ESG performance was subject to mandatory ESG disclosure requirements; zero otherwise (*Mandatory_Disclosure*).

We first confirm that these mandatory ESG disclosure requirements did indeed improve firms' ESG disclosures. Because the firms in our sample tend to be quite large, they may have already been voluntarily disclosing ESG information at the level required by these mandatory requirements. Thus it is possible that these mandatory disclosure requirements may not have resulted in an increase in ESG disclosure. Therefore, we perform a validation analysis before performing our difference-in-differences analysis. We do this by regressing *ESG_Disclosure* on *Mandatory_Disclosure*, with the same control variables as before, along with firm and year fixed effects. As shown in Table 6, the estimated coefficient on *Mandatory_Disclosure* is positive and statistically significant, confirming that the disclosure requirements improved firm's ESG disclosures, as expected.

We then perform a difference-in-differences analysis with firm and year fixed effects, where we re-estimate Equation 1 and replace *ESG_Disclosure* with *Mandatory_Disclosure*. As reported in Table 6, we observe that *Mandatory_Disclosure* is positive and statistically significant, consistent with our primary results.

To address the parallel trends assumption in this analysis, we also perform a dynamic difference-in-differences analysis. In this analysis, we re-estimate our models and replace the

³² For further details on these disclosure requirements, see <https://www.carrotsandsticks.net>. Also note that our tests capture the first instance of broad ESG disclosure requirements in a country. Some countries have subsequently issued additional legislation that requires even greater ESG disclosure than the original mandate.

Mandatory Disclosure dummy with indicator variables for different periods around the adoption of mandatory ESG regulations. For example, *Mandatory Disclosure* $t = -1$ is an indicator variable equal to one in the year prior to when a firm became subject to mandatory ESG disclosure. As reported in Table 6, we see that the coefficients on *Mandatory Disclosure* in the pre-period years are statistically insignificant, suggesting that the parallel trends assumption holds. We also observe in the dynamic validation analysis that the mandatory disclosure rules appear to have increased ESG disclosure around the time that the laws and regulations went into effect. Importantly, we also observe that the ESG Disagreement, on average, appears to increase around this time, although with a delay. Collectively, these results suggest that the findings from our difference-in-differences test are not due to a pre-existing trend.³³

These findings provide further support for the notion that greater ESG disclosure leads to greater ESG disagreement. That being said, a few caveats are worth highlighting. First, many of our observations that take the value of one for *Mandatory_Disclosure* come from a relatively small number of countries thereby making these results dependent on these countries. Second, some of the countries, despite having disclosure mandates, have historically had very low levels of disclosure suggesting that not all such mandates have been effective (e.g., Greece).³⁴ With those caveats in mind, we believe that these tests need to be interpreted with caution, but are additive to the mosaic of evidence we present in the paper.

Inputs vs. Outcomes: Disagreement at the Metric Level

To shed more light on what ESG ratings capture and how ESG disclosure contributes to

³³ We also perform placebo tests where we randomly assign the years that treatment firms became subject to mandatory disclosure requirements. We did this 1000 different times and found that in 98 percent of the regressions there was *not* a positive and significant association between the placebo *Mandatory Disclosure* variable and *ESG_Disagreement* (untabulated). These findings provide additional confidence in the reliability of the findings from our mandatory disclosure tests reported Table 6.

³⁴ In untabulated analyses we also find that the increase in disagreement after the mandatory disclosure requirements is primarily driven by UK firms. Nevertheless, UK firms do not drive the main results reported in Table 4.

ESG rating disagreement, we next examine the individual metrics underlying ESG ratings. When ESG rating agencies evaluate a company's ESG performance to assign a rating, they accumulate and analyze a host of metrics. Some of these metrics are inputs, and others are outcomes. Inputs are the efforts that a company makes to achieve a desired outcome. Outcomes are the performance outcomes resulting from a company's efforts. For example, the presence of a diversity policy is an input, whereas the percentage of women at the firm is an outcome of that policy. Similarly, setting a carbon emissions reduction target is an input, while the carbon intensity of a business is the outcome. An employee health and safety strategy is an input, while lost time injury rates is the outcome. These two sets of metrics are not only conceptually different but they also behave differently, as we show hereafter.

Two of the three ESG ratings agencies that we examine in this paper, Sustainalytics and Thomson Reuters, provide firm-level data on the individual metrics they use to come up with their overall ratings.³⁵ For these raters we classify each of the metrics they use as inputs or outcomes, and then map each metric of Sustainalytics to its corresponding metric(s) from Thomson Reuters based on the theme the metric is meant to reflect.³⁶ This generates 927,352 firm-year-metric matches where we can examine ratings disagreement at the metric level.

This mapping leads to three categories of metrics. First, one where both raters use input metrics. For example, they both may assess a company's policies or initiatives to increase workforce diversity. Second, both raters assess outcome metrics, such as the percentage of women managers. Third, one data provider assesses an input metric while the other assesses an outcome

³⁵ MSCI does not provide their scoring on individual metrics for each firm.

³⁶ There are some topics that are covered by one rating agency and not the other. In these cases, no matching metric exists and we cannot calculate a disagreement score. Thus, we exclude these metrics from our analysis. Additionally, as the focus of our paper is on ESG disclosure, we also exclude metrics that capture controversies identified in the media, as these are typically not identified by firms' disclosures.

metric. For example, while both rating agencies may focus on employee health and safety, one agency focuses on the strength of the company's initiatives to prevent accidents and illnesses (an input) and the other focuses on the total number of injuries and fatalities (an outcome).³⁷

Our framework allows us to both predict for which metric category the level of disagreement will be higher and for which category disclosure might exacerbate disagreement more, thereby increasing our understanding of ESG ratings as a construct. First, based on our theory pointing to a lack of an institutionalized and shared understanding of the meaning of ESG metrics, we expect the lowest disagreement when both raters are assessing input metrics as it is much easier to agree on whether the effort is taking place. Nevertheless, disagreement may still exist for this category because different data providers might disagree on the extent of that effort. For example, data providers likely evaluate not only the presence of a policy but also how strong that policy is. Second, we predict that disagreement will be even higher when both raters evaluate outcome metrics because it is not clear what is a good or a bad outcome. Making that evaluation requires a shared understanding of how to evaluate whether the intention behind the efforts has been achieved. Third, we predict the highest disagreement when one data provider evaluates an input and the other evaluates an outcome, as comparing efforts and outcomes can be like comparing apples and oranges.

To test these predictions, we examine ESG disagreement at the individual metric level using the following OLS regression model:

$$\begin{aligned} ESG\ Metric\ Disagreement_{it} &= \beta_0 + \beta_1 Both\ Outcomes_{it} + \beta_2 Input/Outcome_{it} \\ &+ \sum \beta_k Controls + \varepsilon_{i,t} \end{aligned} \quad (2)$$

³⁷ From a descriptive perspective (untabulated), it appears that ESG raters heavily utilize input metrics, as 81 percent of the observations in our matched-metric sample relate to instances where both raters are using input metrics. In contrast, outcome metrics appear to be far less common, as only 10 percent of the observations pertain to instances where both raters are using outcome metrics. The remaining 9 percent of the observations pertain to input/outcome matches.

where *ESG Metric Disagreement* is the standard deviation of the Sustainalytics rating and Thomson Reuters rating for the matched metrics.³⁸ Our variables of interest are indicator variables that flag each of the three combinations of inputs and outcomes. As we expect the least disagreement when both raters evaluate inputs (*Both Inputs*), we exclude this category from the model and use it as the baseline category. *Both Outcomes* is an indicator variable set equal to one where both rating agencies evaluate outcome metrics. *Input/Outcome* is an indicator variable set equal to one where one rating agency assesses an input and the other an outcome metric. *Controls* captures the same firm-level characteristics as in our prior tests. We expect $\beta_1 > 0$ and $\beta_2 > \beta_1$.

As predicted, in columns (1) and (2) of Table 7 Panel A we find that the estimated coefficients on *Input/Outcome* and *Both Outcomes* are both positive and significant. Further, the estimated coefficient on *Input/Outcome* is statistically larger than the coefficient on *Both Outcomes* ($p < 0.01$, untabulated), as expected.

Having established these baseline predictions, we next consider the role of disclosure in this process. We expect that disagreement across all the three categories should increase as a function of disclosure because disclosure provides more pieces of information to evaluate thereby leading to more disagreement. To test this, we add an indicator for above median ESG Disclosure (*High ESG Disclosure*) to Equation 2 and interact it with each of the metric categories. As reported in columns (3) and (4) of Table 7 Panel A, we find that the coefficient on *High ESG Disclosure* is positive and significant, consistent with our primary findings that greater disclosure is associated with greater disagreement. More interestingly, we find that the effect of disclosure on disagreement is the largest for the category when both data providers evaluate outcomes (i.e., *High ESG*

³⁸ In other words, *ESG Metric Disagreement* is measured the same way *ESG Disagreement* is, but now at the metric level, rather than at the firm level. The results are similar if we instead measure ESG metric disagreement by taking the absolute value of the difference in the metric scores (untabulated).

Disclosure × *Both Outcomes* is positive and significant). Again, this is consistent with our theory that the evaluation of outcomes relies on having an institutionalized understanding of what a good vs. a bad outcome might be. Thereby, we find a stronger effect on disagreement for the category of metrics that one would expect the largest need for development of a shared understanding of what the metrics might mean.

These latter results allow us to make some inferences about the ‘desirability’ of the disagreement. While we cannot pursue a general equilibrium evaluation of the social welfare implications from ESG disagreement, these findings suggest that to improve accountability among firms there needs to be heavy emphasis on outcomes. In the absence of focusing on outcome metrics we could have extended ‘goodwashing’ among companies. Yet our results also provide evidence that getting agreement on outcomes is much more difficult and that disagreement increases sharply with more disclosure. In this context, we regard disagreement as inhibiting the accountability process and that disclosure needs to be shaped around a consistent set of rules and a framework that allows analysts to evaluate those outcome metrics with clear benchmarks.

To provide further insight into this phenomenon, we also examine the role of ESG disclosure at the industry level. A high degree of ESG disclosure at the industry level might help create a common benchmark against which rating agencies can assess a given firm’s ESG performance, thereby decreasing disagreement. However, due to the highly subjective nature of ESG information and the lack of common norms for evaluating ESG information, we predict that the opposite would occur, as greater industry disclosure likely provides even more complex information for raters to try to interpret, and even more information to disagree about. To test this, we create an indicator variable set to one if the average level of ESG disclosure for a firm’s industry in year t is above the sample median industry ESG disclosure (*Industry High ESG Disclosure*),

and replace our firm-level ESG disclosure variable with this measure.³⁹ Consistent with our prediction, in columns 1 and 2 of Table 7 Panel B we find that greater industry-wide ESG disclosure appears to exacerbate disagreement for outcome metrics.⁴⁰ This finding holds in columns 3 and 4 when we add back firm level ESG disclosure to the model.⁴¹ Overall, these findings suggest that widespread increases in ESG disclosure across firms is unlikely to resolve ESG disagreement until a shared understanding is developed regarding what constitutes good and bad ESG performance.

Consequences of ESG Disagreement

To better understand how ESG rating disagreement influences markets, we next examine stock market consequences of ESG disagreement. Specifically, we perform short-window tests using 3 day windows (-1,+1) centered on the date when a rating agency comes out with a new ESG rating. These tests allow us to understand whether the issuance of a new rating that disagrees more with the existing rating from another agency has implications for market prices. In these tests we use data from MSCI and Sustainalytics as their data reveal the exact day when the new rating comes out, and their ratings are released in a staggered fashion throughout the year. (In contrast, Thomson Reuters only provides the year their annual ratings pertains to.)⁴² We perform these market consequence tests by estimating the following OLS regression model:

³⁹ Industries such as chemicals, tobacco, and electrical equipment tend to have high ESG disclosure, whereas industries like personal services, entertainment, and trading tend to have low ESG disclosure (untabulated).

⁴⁰ We speculate that the negative and significant coefficient on the base term for *Both Outcomes* in Panel B is due to the fact that in industries with low ESG disclosure raters might rely more on inputs in how they rate companies as evaluations of outcomes might be even more difficult to perform.

⁴¹ Moreover, in untabulated analyses, when we examine the combination of high firm-level ESG disclosure with high industry-level ESG disclosure, we find that the effect of high firm-level ESG disclosure on ESG outcome disagreement manifests in industries with greater ESG disclosure.

⁴² Sustainalytics and MSCI often update their ratings more than once a year. We exclude observations where rating agencies issue updated ratings that do not change their prior rating (i.e., rating reiterations), as prior research on other information intermediaries has found dampened market reactions to reiterations (e.g., Mikhail, Walther, and Willis 2004). If we also include rating reiterations in these tests, we observe similar, although economically weaker, results as expected (untabulated). Also note that the sample for these consequences tests contains ratings released up through the end of 2017, consistent with our determinants tests.

$$\text{Market Outcome}_{i(t-1,t+1)} = \beta_0 + \beta_1 \text{ESG_Disagreement}_{it} + \sum \beta_k \text{Controls} + \varepsilon_{i,t} \quad (3)$$

where *Market Outcome* is measured using three different variables: *Absolute CAR*, *Return Volatility*, and *Bid-Ask Spread*. *Absolute CAR* captures absolute market-adjusted returns cumulated over the announcement window. We evaluate unsigned instead of signed returns as disagreement is unsigned and therefore we do not have ex ante a view about whether disagreement would represent good or bad news from an equity holder perspective. *Return Volatility* is the standard deviation of market-adjusted returns during the announcement window; and *Bid-Ask Spread* is the average daily bid-ask spread during the announcement window. *ESG Disagreement* is measured the same as before, except now it is based on the new ESG rating released on day t (e.g., from MSCI) and the most recently available ESG rating from the other rater (e.g., Sustainalytics). *Controls* includes *ESG_Avg*, which is also measured based on the most recently available ESG ratings to help control for any new economic events that may have occurred since the last rating update. The other controls are the same variables as in our prior tests, measured as of the most recent fiscal year-end prior to the rating release date. The models are estimated with (1) industry, nation, and year fixed effects, or (2) firm and year fixed effects.

Table 8 Panel A reports results of these tests. As shown in columns (1) and (2), we find that *ESG Disagreement* is positively associated with *Absolute CAR*. In economic terms, increasing *ESG Disagreement* from the 25th to the 75th percentile is associated with an eight to ten basis point increase in absolute market-adjusted returns, depending on the specification. Given that the sample is comprised of very large firms, this effect size seems reasonable. In columns (3) and (4) we find that *ESG Disagreement* is also positively associated with *Return Volatility*. Here, increasing *ESG Disagreement* from the 25th to the 75th percentile is associated with a three to four basis point increase in return volatility. Lastly, when we examine bid-ask spreads in columns (5) and (6), we

do not observe a significant relation with ESG Disagreement. Collectively, these findings that ESG disagreement is positively associated with return volatility and absolute price movements suggests that ESG disagreement is relevant to market participants and influences stock prices.⁴³

As the use of ESG data has been increasing over time (Gilbert 2019), we next examine whether these market consequences are stronger in the later part of our sample. To test this, we interact a *Time* trend variable with *ESG Disagreement* and re-run these analyses.⁴⁴ As reported in Table 8 Panel B, we find that the interaction term (*ESG Disagreement* \times *Time*) is positive and significant in one of the two specifications when *Absolute CAR* is the dependent variable, and in both specifications when *Return Volatility* is the dependent variable. These results provide some evidence that ESG disagreement is having an increasing impact on markets as time goes on.

Lastly, to shed more light on the capital market implications of ESG disagreement, we explore the influence of ESG disagreement on firms' financing choices. Prior research has found that greater ESG performance tends to increase firms' access to external capital (Cheng, Ioannou, and Serafeim 2014). As our prior findings illustrate that ESG disagreement appears to introduce uncertainty regarding the sustainability of a firm, this may serve as a market friction that impedes a firm's access to external capital. Thus, firms facing greater ESG disagreement may be less likely to obtain external financing and instead may rely more on internal financing, as internal financing tends to be less costly. To test this, we utilize our firm-year dataset and examine the following three dependent variables: *Equity Issuance*, *Debt Issuance*, and *Cash Holdings*. *Equity Issuance* is an indicator variable equal to one if the firm issues equity in year $t+1$; *Debt Issuance* is an indicator variable equal to one if the firm issues debt in year $t+1$; *Cash Holdings* is cash divided by total

⁴³ The results are similar if we also drop (or control for) observations where firms announced earnings or released management guidance during the 3 day window (untabulated).

⁴⁴ The main effect of the time trend variable is subsumed by the year fixed effects and thus is excluded in the model.

assets in year $t+1$. We predict that *ESG Disagreement* will be negatively associated with the two external financing measures (*Equity Issuance* and *Debt Issuance*) and positively associated with the internal financing measure (*Cash Holdings*).

As reported in Table 9 Panel A, the results from OLS regressions generally support these predictions, as we find statistically significant results in the predicted directions in four of the six specifications. These findings suggest that firms facing greater ESG disagreement are less likely to raise external financing and instead tend to rely more on internal financing.⁴⁵ In economic terms, increasing *ESG Disagreement* from the 25th to the 75th percentile is associated with a 0.2 to 1.1 (0 to 2.2) percent lower likelihood of issuing equity (debt) in year $t+1$, depending on the model specification. Similarly, increasing *ESG Disagreement* from the 25th to the 75th percentile is associated with a 1.4 to 5.1 percent increase in a firm's cash holdings in year $t+1$, depending on the model. Additionally, in Panel B, when we perform time trend tests for these financing outcomes, we find some evidence that ESG disagreement is having an increasing impact on these financing outcomes as time goes on.

VII. CONCLUSION

The use of environmental, social, and governance (ESG) ratings in financial markets has risen dramatically, and so has the disagreement across ESG rating agencies regarding what rating to give to an individual firm. The media has highlighted the existence of ESG rating disagreement, but there is a dearth of evidence on what drives this disagreement. In this paper, we document empirical evidence on the extent to which a firm's ESG disclosure helps explain this disagreement.

Our tests reveal several interesting findings. In contrast to evidence in other settings where greater disclosure helps reduce disagreement among information intermediaries, we find that

⁴⁵ For the binary dependent variables, we find similar results if we instead use probit models (untabulated).

greater ESG disclosure leads to greater ESG disagreement across ESG rating agencies. These findings are robust to including firm fixed effects and using a difference-in-differences design with staggered mandatory ESG disclosure shocks. These findings also appear to be primarily driven by the environmental and social disclosures, rather than governance disclosures. We also find that ESG disclosure appears to amplify disagreement about ESG metrics, particularly for ESG outcomes. Overall, our results show that greater ESG disclosure does not appear to help resolve ESG rating disagreement.

We note that over time as analysts develop a consensus both on the metrics to use to assess a firm's performance on a specific ESG issue and how to interpret the information reflected in each metric, the relation between disclosure and disagreement might diminish or even become negative. In other words, our study is likely to be reflective of the early stages of institutional innovation around ESG disclosures.

As rating disagreements threaten to dampen confidence in ESG ratings, the findings in our study are important because without understanding why providers disagree, it becomes difficult to understand not only what the potential remedies could be, but also the plausible consequences of this disagreement. Although we also document some of the consequences of ESG disagreement (e.g., higher return volatility and larger absolute price movements), there is still much to learn about the consequences of disagreement, which we encourage future research to explore.

REFERENCES

- Akins, B. 2018. Financial reporting quality and uncertainty about credit risk among ratings agencies. *The Accounting Review* 93(4): 1-22.
- Allen, K. 2018. Lies, damned lies and ESG rating methodologies. *Financial Times*. Dec. 6. <https://ftalphaville.ft.com/2018/12/06/1544076001000/Lies--damned-lies-and-ESG-rating-methodologies/>
- Baldini, M., L. Dal Maso, G. Liberatore, F. Mazzi, and S. Terzani. 2018. Role of country-and firm-level determinants in environmental, social, and governance disclosure. *Journal of Business Ethics* 150(1): 79-98.
- Bamber, L., O. Barron, and D. Stevens. 2011. Trading volume around earnings announcements and other financial reports: Theory, research design, empirical evidence, and directions for future research. *Contemporary Accounting Research* 28(2): 431-471.
- Bamber, L., O. Barron, and T. Stober. 1999. Differential interpretations and trading volume. *Journal of Financial and Quantitative Analysis* 34(3): 369-386.
- Barinov, A. 2013. Analyst disagreement and aggregate volatility risk. *Journal of Financial and Quantitative Analysis* 48(6): 1877-1900.
- Becker, H. 1982. *Art Worlds*. Berkeley: Univ. Calif. Press.
- Berg, F., J. Koelbel, and R. Rigobon. 2020. Aggregate confusion: The divergence of ESG ratings. Working paper.
- Bereskin, F., S.K. Byun, M.S. Officer, and J.M. Oh. 2018. The effect of cultural similarity on mergers and acquisitions: Evidence from corporate social responsibility. *Journal of Financial and Quantitative Analysis* 53(5): 1995-2039.
- Bloomfield, R. and P. Fischer. 2011. Disagreement and the cost of capital. *Journal of Accounting Research* 49(1): 41-68.
- Bonsall, S., K. Koharki, and L. Watson. 2017. Deciphering tax avoidance: Evidence from credit rating disagreements. *Contemporary Accounting Research* 34(2): 818-848.
- Bonsall, S. and B. Miller. 2017. The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies* 22(2): 608-643.
- Brackley, A., E. Petroy, and C. Wong. 2019. Rate the raters 2019: Expert survey results. SustainAbility, an ERM Group Company.
- Cantor, R. and F. Packer. 1994. The credit rating industry. *Quarterly Review* 19(2): 1-26.
- Chatterji, A., R. Durand, D. Levine, and S. Touboul. 2016. Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal* 37(8): 1597-1614.
- Chatterji, A. and M. Toffel. 2010. How firms respond to being rated. *Strategic Management Journal* 31(9): 917-945.
- Cheng, B., I. Ioannou, and G. Serafeim. 2014. Corporate social responsibility and access to finance. *Strategic Management Journal* 35(1): 1-23.
- Christensen, D.M. 2016. Corporate accountability reporting and high-profile misconduct. *The Accounting Review* 91(2): 377-399.
- Cookson, J.A, and M. Niessner. 2020. Why don't we agree? Evidence from a social network of investors. *Journal of Finance* 75(1): 173-228.
- Daines, R., I. Gow, and D. Larcker. 2010. Rating the ratings: How good are commercial governance ratings? *Journal of Financial Economics* 98(3): 439-461.

- Delmas, M.A. and V.C. Burbano. 2011. The drivers of greenwashing. *California Management Review* 54(1): 64-87.
- Espeland, W.N. and M.L. Stevens. 1998. Commensuration as a social process. *Annual Review of Sociology* 24(1): 313-343.
- Ferrell, A., H. Liang, and L. Renneboog. 2016. Socially responsible firms. *Journal of Financial Economics* 122(3): 585-606.
- Gilbert, M. 2019. The rising cost of investing responsibly. *Bloomberg*. Jan. 10. <https://www.bloomberg.com/opinion/articles/2019-01-11/the-rising-cost-of-esg-and-socially-responsible-investing>
- Global Sustainable Investment Alliance (GSIA). 2018. *Global Sustainable Investment Review*. http://www.gsi-alliance.org/wp-content/uploads/2019/03/GSIR_Review2018.3.28.pdf.
- Grewal, J. and G. Serafeim. 2020. Research on corporate sustainability: review and directions for future research. *Foundations and Trends in Accounting* 14(2): 73–127.
- Guest, P. and M. Nerino. 2020. Do corporate governance ratings change investor expectations? Evidence from announcements by Institutional Shareholder Services. *Review of Finance* 24(4): 891-928.
- Hirshleifer, D. and S.H. Teoh. 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36(1–3): 337–386.
- Hope, O.K. 2003. Accounting policy disclosures and analysts' forecasts. *Contemporary Accounting Research* 20(2): 295-321.
- Hubbard, T.D., D.M. Christensen, and S.D. Graffin. 2017. Higher highs and lower lows: The role of corporate social responsibility in CEO dismissal. *Strategic Management Journal* 38(11): 2255-2265.
- Isaac, A., and B. Dummett. 2020. London Stock Exchange Group sells Italian exchange for \$5.1 billion, paving way for Refinitiv deal. *Wall Street Journal Online*. October 9.
- Jones, E., L. Rock, K. Shaver, G. Goethals, and L. Ward. 1968. Pattern of performance and ability attribution: An unexpected primacy effect. *Journal of Personality and Social Psychology* 10(4): 317-340.
- Kennedy, P. 2008. A guide to econometrics. Sixth edition. Malden, MA: Blackwell Publishing.
- Khan, M., G. Serafeim, and A. Yoon. 2016. Corporate sustainability: First evidence on materiality. *The Accounting Review* 91(6): 1697-1724.
- Kimbrough, M., X. Wang and S. Wei. 2019. Does voluntary disclosure resolve disagreement outside of the capital market setting? Evidence from corporate social responsibility reporting. Working paper.
- Knorr-Cetina, K. 1999. Epistemic cultures: The cultures of knowledge societies. Cambridge, MA: Harvard.
- Kotsantonis, S. and G. Serafeim. 2019. Four things no one will tell you about ESG data. *Journal of Applied Corporate Finance* 31(2): 50-58.
- Krüger, F. and I. Nolte. 2016. Disagreement versus uncertainty: Evidence from distribution forecasts. *Journal of Banking & Finance*, 72: S172-S186.
- Lamont, M. 2009. *How professors think*. Harvard University Press.
- Lamont, M. 2012. Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology* 38: 201-221.
- Lang, M. and R. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71(4): 467-492.

- Li, Y., M. Gong, X.Y. Zhang, and L. Koh. 2018. The impact of environmental, social, and governance disclosure on firm value: The role of CEO power. *The British Accounting Review* 50(1): 60-75.
- Merton, R.C. 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42(3): 483–510.
- Mikhail, M., B. Walther, and R. Willis. 2004. Do security analysts exhibit persistent differences in stock picking ability? *Journal of Financial Economics* 74(1): 67-91.
- Morgan, D.P. 2002. Rating banks: Risk and uncertainty in an opaque industry. *American Economic Review* 92(4): 874-888.
- MSCI. 2018. MSCI ESG ratings methodology. Available at: <https://www.msci.com>
- Sadka, R. and A. Scherbina. 2007. Analyst disagreement, mispricing, and liquidity. *Journal of Finance* 62(5): 2367-2403.
- Sindreu, J. and S. Kent. 2018. Why it's so hard to be an 'ethical' investor. *The Wall Street Journal Online*. Sept. 1. <https://www.wsj.com/articles/why-its-so-hard-to-be-an-ethical-investor-1535799601>
- Sørensen, J.B. 2002. The use and misuse of the coefficient of variation in organizational demography research. *Sociological Methods & Research* 30(4): 475-491.
- Sustainalytics. 2018. Sustainalytics' ESG rating research methodology. Available at: <https://www.sustainalytics.com>
- Thomson Reuters. 2017. Thomson Reuters ESG scores. March. Available at: <https://www.thomsonreuters.com>
- Thomson Reuters. 2018. Thomson Reuters and Blackstone close Financial & Risk transaction. Available at: <https://www.thomsonreuters.com/en/press-releases/2018/october/thomson-reuters-and-blackstone-close-financial-and-risk-transaction.html>

APPENDIX A
Variable Definitions

Variable Name	Description	Source
ESG Variables		
<i>ESG_Score</i>	The firm's overall weighted environmental, social, and governance rating from a unique ESG rating agency for year <i>t</i> on a scale of 0 to 100.	MSCI IVA, TR ASSET4, Sustainalytics
<i>E_Score</i>	The firm's environmental rating from a unique ESG rating agency for year <i>t</i> on a scale of 0 to 100.	MSCI IVA, TR ASSET4, Sustainalytics
<i>S_Score</i>	The firm's social rating from a unique ESG rating agency for year <i>t</i> on a scale of 0 to 100.	MSCI IVA, TR ASSET4, Sustainalytics
<i>G_Score</i>	The firm's governance rating from a unique ESG rating agency for year <i>t</i> on a scale of 0 to 100.	MSCI IVA, TR ASSET4, Sustainalytics
<i>ESG_Avg</i>	The average environmental, social, and governance rating a firm received for year <i>t</i> 's ESG performance.	MSCI IVA, TR ASSET4, Sustainalytics
<i>E_Avg</i>	The average environmental rating a firm received for year <i>t</i> 's environmental performance.	MSCI IVA, TR ASSET4, Sustainalytics
<i>S_Avg</i>	The average social rating a firm received for year <i>t</i> 's social performance.	MSCI IVA, TR ASSET4, Sustainalytics
<i>G_Avg</i>	The average governance rating a firm received for year <i>t</i> 's governance performance.	MSCI IVA, TR ASSET4, Sustainalytics
<i>ESG_Disagreement</i>	The standard deviation of ESG ratings that a firm received for year <i>t</i> 's ESG performance.	MSCI IVA, TR ASSET4, Sustainalytics
<i>E_Disagreement</i>	The standard deviation of environmental ratings that a firm received for year <i>t</i> 's environmental performance.	MSCI IVA, TR ASSET4, Sustainalytics
<i>S_Disagreement</i>	The standard deviation of social ratings that a firm received for year <i>t</i> 's social performance.	MSCI IVA, TR ASSET4, Sustainalytics
<i>G_Disagreement</i>	The standard deviation of governance ratings that a firm received for year <i>t</i> 's governance performance.	MSCI IVA, TR ASSET4, Sustainalytics
<i>Diff_TR_MSCI</i>	The absolute value of the difference between the rating Thomson Reuters and MSCI gave a firm for year <i>t</i> 's ESG performance.	MSCI IVA, TR ASSET4, Sustainalytics
<i>Diff_TR_SUST</i>	The absolute value of the difference between the rating Thomson Reuters and Sustainalytics gave a firm for year <i>t</i> 's ESG performance.	MSCI IVA, TR ASSET4, Sustainalytics

APPENDIX A (Continued)
Variable Definitions

Variable Name	Description	Source
<i>Diff_MSCI_SUST</i>	The absolute value of the difference between the rating MSCI and Sustainalytics gave a firm for year <i>t</i> 's ESG performance.	MSCI IVA, TR ASSET4, Sustainalytics
<i>ESG_Disclosure</i>	The firm's ESG disclosure score for the ESG report pertaining to year <i>t</i> 's performance.	Bloomberg
<i>E_Disclosure</i>	The firm's environmental disclosure score for the ESG report pertaining to year <i>t</i> 's performance.	Bloomberg
<i>S_Disclosure</i>	The firm's social disclosure score for the ESG report pertaining to year <i>t</i> 's performance.	Bloomberg
<i>G_Disclosure</i>	The firm's governance disclosure score for the ESG report pertaining to year <i>t</i> 's performance.	Bloomberg
<i>High_ESG_Disclosure</i>	An indicator variable equal to one if the firm-year's ESG disclosure was above the sample median; zero otherwise.	Bloomberg
<i>Industry High ESG Disclosure</i>	An indicator variable equal to one if the average ESG disclosure in the firm's industry during year <i>t</i> was above the sample median industry disclosure; zero otherwise.	Bloomberg, Worldscope
<i>Mandatory_Disclosure</i>	An indicator variable equal to one if the firm-year's ESG performance was subject to broad mandatory ESG disclosure requirements; zero otherwise.	Carrots & Sticks
<i>ESG_Metric_Disagree</i>	The standard deviation of scores that a firm received for year <i>t</i> 's performance for the matched metrics.	TR ASSET4, Sustainalytics
Control Variables		
<i>Firm Size</i>	The natural log of total assets (in millions of US\$) as of the end of year <i>t</i> .	Worldscope
<i>ROA</i>	Return on Assets, defined as net income for year <i>t</i> divided by total assets at the end of year <i>t</i> .	Worldscope
<i>BTM</i>	The book value of equity divided by the market value of equity, as of the end of year <i>t</i> .	Worldscope
<i>Leverage</i>	Total liabilities divided by total assets, as of the end of year <i>t</i> .	Worldscope
<i>Analyst Following</i>	The natural log of the number of analysts following the firm as of the end of year <i>t</i> .	Worldscope
<i>Inst. Ownership</i>	The percentage of the firm's shares owned by institutional investors at the end of year <i>t</i> , multiplied by 100. Set to zero if missing. Max value capped at 100.	Factset
<i>Industry Fixed Effects</i>	Industry indicators are based on the Fama French 49 industry classifications.	Ken French's website

APPENDIX A (Continued)
Variable Definitions

Variable Name	Description	Source
<i>Time</i>	The fiscal year for a particular firm-year observation minus the first year in the sample.	Worldscope
Input/Outcome Variables		
<i>Both Inputs</i>	An indicator equal to 1 if both ESG raters measure the issue using input metrics.	TR Asset4, Sustainalytics
<i>Both Outcomes</i>	An indicator equal to 1 if both ESG raters measure the issue using outcome metrics.	TR Asset4, Sustainalytics
<i>Input/Outcome</i>	An indicator equal to 1 if one ESG rater measures the issue using an input metric, and the other rater uses an outcome metric.	TR Asset4, Sustainalytics
Consequence Variables		
<i>Absolute CAR</i>	The firm's cumulative absolute market-adjusted return during the 3-day window (-1,+1) around the ESG rating release date, multiplied by 100 for ease of exposition.	Datastream
<i>Return Volatility</i>	The standard deviation of the firm's market-adjusted returns during the 3-day window (-1,+1) around the ESG rating release date, multiplied by 100 for ease of exposition.	Datastream
<i>Bid-Ask Spread</i>	The average daily relative bid-ask spread during the 3-day window (-1,+1) around the ESG rating release date, multiplied by 100 for ease of exposition. The relative bid-ask spread is measured as $(Ask-Bid)/((Ask+Bid)/2)$.	Datastream
<i>Equity Issuance</i>	An indicator equal to 1 if proceeds from equity issuances in year $t+1$ are greater than zero.	Worldscope
<i>Debt Issuance</i>	An indicator equal to 1 if net proceeds from debt issuances in year $t+1$ are greater than zero.	Worldscope
<i>Cash Holdings</i>	Cash, scaled by total assets at time $t+1$.	Worldscope

APPENDIX B
Mandatory ESG Disclosure Requirements (by Country)

Country	Source of Requirement	Year	Regulation
Portugal	Government	1985	Law No 141/85 and Decree-Law No 9/92
Australia	Stock Exchange	1996	Australian Stock Exchange Listing Rule 4.10.3
Finland	Government	1997	The Finnish Accounting Act
France	Government	2003	New Economic Regulations Act
Greece	Government	2007	Law 3487
Malaysia	Stock exchange	2007	Bursa Malaysia Listing Requirement (Appendix 9C, Part A, 29)
Argentina	Government	2008	Law 2594
China	Stock exchange	2008	Guidelines for Corporate Social Responsibility of Shenzhen Stock Exchange Listed Companies; Shanghai Stock Exchange Guidelines on Listed Companies' Environmental Information Disclosure
Sweden	Government	2008	Guidelines for external reporting by state-owned companies
Austria	Government	2009	Corporate Law Amendment Act 2008 - URÄG 2008
Denmark	Government	2009	Act amending the Danish Financial Statements Act. (Report on social responsibility for large businesses).
Netherlands	Government	2010	Guideline 400 (RJ-Uiting 2009-8) and Parliamentary Paper 26485, no.86
Pakistan	Government	2010	S.R.O. 983(I)/2009 Companies (Corporate Social Responsibility) General Order, 2009
South Africa	Stock exchange	2010	Johannesburg Stock Exchange Listing Requirement 8.63 and King Code of Governance Principles (King III)
India	Government	2012	Securities and Exchange Board CIR/CFD/DIL/8/2012
Indonesia	Government	2012	Decision of the Chairman of the Capital Markets Supervisory Agency No.KEP-431/BL/2012
Israel	Government	2012	Public Reporting Directives [1] (10/11) Section 620-E

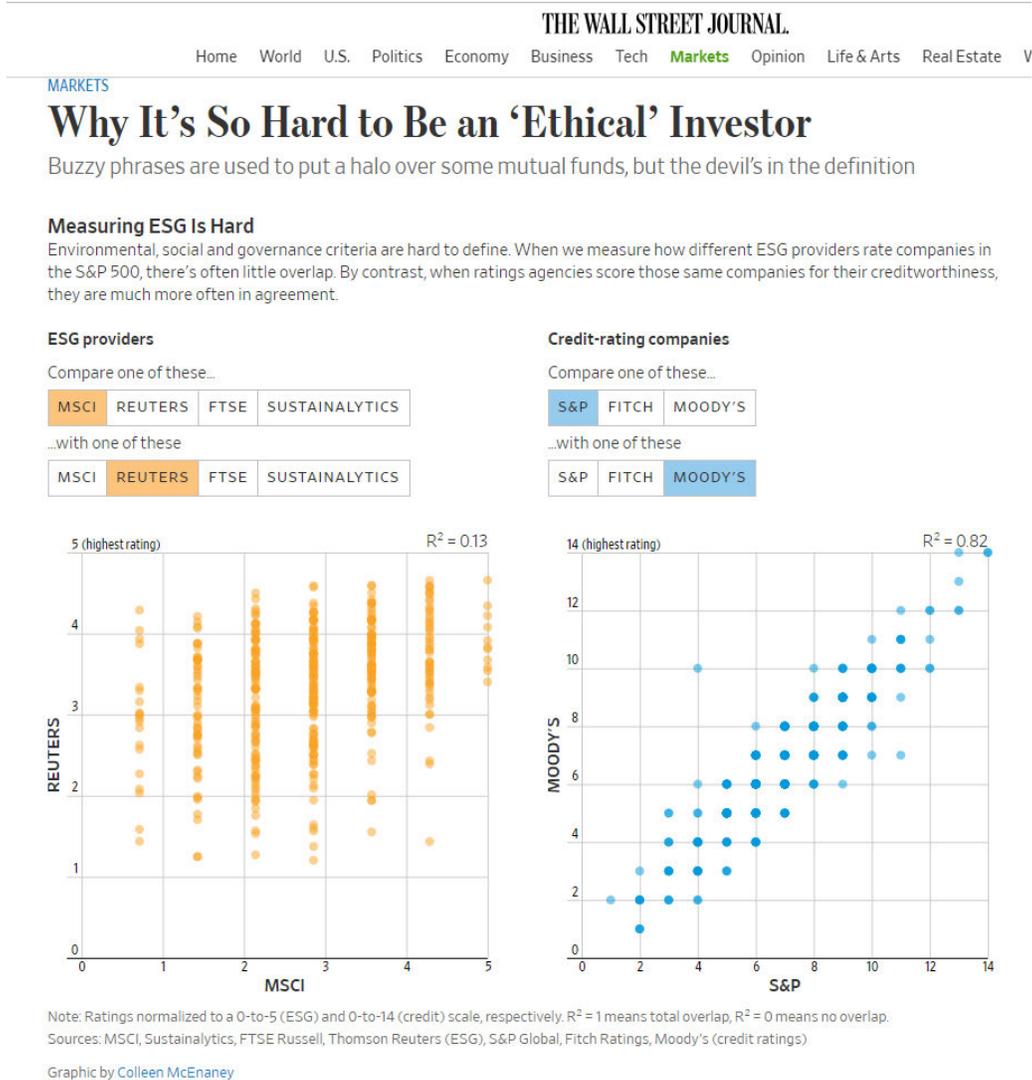
APPENDIX B (Continued)
Mandatory ESG Disclosure Requirements (by Country)

Country	Source of Requirement	Year	Regulation
Nigeria	Government	2012	Nigerian Sustainable Banking Principles
Norway	Government	2013	Statute 44; LOV-2013-04-19-15
Qatar	Government	2013	Sustainable Development Industry Reporting (SDIR) Program
United Kingdom	Government	2013	UK Statutory Instruments 2013 No. 1970
Taiwan	Stock exchange	2014	Taiwan Stock Exchange Corporation Rules Governing the Preparation and Filing of Corporate Social Responsibility Reports by TWSE Listed Companies
Thailand	Government	2014	Rules, Conditions and Procedures for Disclosure regarding Financial and Non-financial Information of Securities Issuers
Turkey	Government	2014	Communiqué on Corporate Governance
Chile	Government	2015	General Standard N° 385
Peru	Government	2015	Resolution SMV No 033-2015-SMV/01
Hong Kong	Stock exchange	2016	HKEx Environmental, Social and Governance Reporting Guide

This appendix summarizes staggered adoptions of broad mandatory ESG disclosure requirements through the end of our sample period. Some ESG disclosure requirements are initiated by individual countries, and others are initiated by stock exchanges in those countries. The “year” column captures the first fiscal year-ends that became subject to the disclosure requirements. The “regulation” column captures the name of the regulation identified by Carrots and Sticks and/or by internet searches. During our sample period, in a few countries the disclosure requirements only applied to a subset of firms. The requirements in Israel and Nigeria applied to banks, and in Qatar they applied to oil, energy, and transportation firms. In India they applied to firms on the BSE and National India stock exchanges. In Sweden they applied to state owned enterprises.

APPENDIX C

Although some disagreement in ratings is to be expected when different agencies rate a given firm, the following graphic illustrates how stark the differences can be between ESG raters. Rating disagreement between credit rating agencies is provided as a comparison.



This image comes from Sindreu and Kent (2018). It has been included in this document with permission through a license for re-use of an article and graphics in publication via Copyright Clearance Center. The original interactive graphic is available at: <https://www.wsj.com/articles/why-its-so-hard-to-be-an-ethical-investor-1535799601>.

TABLE 1
Sample Construction

	<u>Firm-Years</u>	<u>Unique Firms</u>
Worldscope firm-years with ESG ratings (2004 - 2016)	65,995	11,217
Require ESG ratings from 2+ raters for each firm-year	<u>(26,892)</u>	<u>(4,268)</u>
	39,103	6,949
Require control variables	<u>(511)</u>	<u>(91)</u>
	38,592	6,858
Require ESG disclosure scores	<u>(7,892)</u>	<u>(1,221)</u>
	30,700	5,637

The table presents the summary of our sample construction. The sample covers the period from 2004 to 2016 for firm-years with at least two ESG ratings from MSCI, Thomson Reuters ASSET4, and Sustainalytics. Control variables and ESG disclosure scores are as described in Appendix A.

TABLE 2
Descriptive Statistics

Panel A: Country Composition

Nation	Freq.	%	Nation	Freq.	%
United States	9,704	31.61	Belgium	172	0.56
Japan	3,444	11.22	Turkey	165	0.54
United Kingdom	2,413	7.86	Bermuda	163	0.53
China	2,187	7.12	Philippines	147	0.48
Canada	1,848	6.02	Poland	136	0.44
Australia	1,561	5.08	Austria	134	0.44
France	852	2.78	Chile	130	0.42
Germany	728	2.37	Israel	97	0.32
Switzerland	561	1.83	New Zealand	97	0.32
India	534	1.74	Greece	90	0.29
South Africa	516	1.68	Portugal	81	0.26
Brazil	481	1.57	Colombia	77	0.25
Sweden	470	1.53	United Arab Emir.	58	0.19
Spain	362	1.18	Luxembourg	56	0.18
Italy	339	1.10	Jersey	46	0.15
Netherlands	332	1.08	Qatar	38	0.12
Singapore	316	1.03	Hungary	28	0.09
Malaysia	266	0.87	Czech Republic	25	0.08
Indonesia	222	0.72	Macau	21	0.07
Finland	218	0.71	Peru	21	0.07
Norway	217	0.71	Egypt	19	0.06
Ireland	215	0.70	Puerto Rico	12	0.04
South Korea	215	0.70	Cayman Islands	11	0.04
Russia	206	0.67	Malta	9	0.03
Mexico	202	0.66	Papua New Guinea	9	0.03
Denmark	198	0.64	Isle of Man	8	0.03
Thailand	187	0.61	Other	56	0.18
				30,700	100.00

TABLE 2 (continued)

Panel B: ESG Ratings from each Rating Agency

	<u>N</u>	<u>Mean</u>	<u>Median</u>	<u>Standard Deviation</u>
MSCI IVA				
<i>ESG Score</i>	26,116	48.36	48.00	12.41
<i>E Score</i>	26,108	50.20	50.00	21.28
<i>S Score</i>	26,116	46.35	46.00	17.87
<i>G Score</i>	26,102	56.13	55.00	20.58
TR ASSET4				
<i>ESG Score</i>	29,767	54.13	56.15	24.53
<i>E Score</i>	29,767	53.67	56.85	32.04
<i>S Score</i>	29,767	54.18	57.21	31.03
<i>G Score</i>	29,767	54.52	60.92	29.58
SUSTAINALYTICS				
<i>ESG Score</i>	25,786	57.21	56.00	9.82
<i>E Score</i>	25,786	54.28	53.00	13.82
<i>S Score</i>	25,786	56.99	56.00	11.06
<i>G Score</i>	25,786	62.21	62.00	10.89

TABLE 2 (continued)

Panel C: Disagreement for each ESG category

	<u>N</u>	<u>Mean</u>	<u>Median</u>	<u>Standard Deviation</u>
ESG Overall				
<i>ESG_Avg</i>	30,700	52.80	52.95	13.93
<i>ESG_Disagreement</i>	30,700	12.32	11.65	6.75
<i>ESG_Diff_TR_MSCI</i>	25,183	19.67	17.90	13.01
<i>ESG_Diff_TR_SUST</i>	24,853	15.75	14.34	10.61
<i>ESG_Diff_MSCI_SUST</i>	21,202	12.71	11.00	9.08
Environmental				
<i>E_Avg</i>	30,700	52.08	52.23	19.11
<i>E_Disagreement</i>	30,700	17.03	16.60	8.38
<i>E_Diff_TR_MSCI</i>	25,175	27.47	26.05	17.29
<i>E_Diff_TR_SUST</i>	24,853	21.37	21.92	11.44
<i>E_Diff_MSCI_SUST</i>	21,194	17.28	15.00	12.28
Social				
<i>S_Avg</i>	30,700	52.08	52.34	16.90
<i>S_Disagreement</i>	30,700	17.45	17.22	8.52
<i>S_Diff_TR_MSCI</i>	25,183	27.10	25.49	17.09
<i>S_Diff_TR_SUST</i>	24,853	22.64	22.05	12.93
<i>S_Diff_MSCI_SUST</i>	21,202	18.06	15.67	13.16
Governance				
<i>G_Avg</i>	30,700	57.12	58.78	15.92
<i>G_Disagreement</i>	30,696	17.53	16.64	9.68
<i>G_Diff_TR_MSCI</i>	25,169	26.74	24.16	18.14
<i>G_Diff_TR_SUST</i>	24,853	23.46	20.62	15.89
<i>G_Diff_MSCI_SUST</i>	21,192	18.59	16.00	13.23

TABLE 2 (continued)

Panel D: Descriptive Statistics of Regression Variables (n = 30,700)

Variables	Mean	Standard Deviation	25%	Median	75%
<i>ESG_Disagreement</i>	12.32	6.75	7.12	11.65	16.92
<i>E_Disagreement</i>	17.03	8.38	11.14	16.60	22.33
<i>S_Disagreement</i>	17.45	8.52	11.19	17.22	23.12
<i>G_Disagreement</i>	17.53	9.68	10.06	16.64	24.09
<i>ESG_Avg</i>	52.80	13.93	41.43	52.95	64.01
<i>E_Avg</i>	52.08	19.11	35.71	52.23	68.36
<i>S_Avg</i>	52.08	16.90	38.09	52.34	65.90
<i>G_Avg</i>	57.12	15.92	45.11	58.78	69.44
<i>ESG_Disclosure</i>	28.50	14.97	14.91	25.62	40.50
<i>E_Disclosure</i>	19.20	18.79	0.00	13.95	34.88
<i>S_Disclosure</i>	26.00	19.32	8.77	23.33	38.60
<i>G_Disclosure</i>	51.78	9.03	48.21	51.79	57.14
<i>Firm Size</i>	8.89	1.64	7.77	8.73	9.87
<i>ROA</i>	0.04	0.08	0.01	0.04	0.08
<i>BTM</i>	0.64	0.50	0.30	0.52	0.85
<i>Leverage</i>	0.58	0.23	0.43	0.58	0.74
<i>Analyst Following</i>	2.43	0.73	2.08	2.56	2.94
<i>Inst. Ownership</i>	48.03	31.28	19.70	41.97	78.26

The table presents the sample descriptive statistics. Panel A shows the sample distribution by country (China includes observations from Taiwan and Hong Kong). Panel B reports summary statistics of ESG ratings from each rating agency. Panel C reports summary statistics of ESG rating disagreement for each ESG category. Panel D reports summary statistics of the control variables. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions.

TABLE 3
Time Trend Analyses

Panel A: ESG rating disagreement over time

Dependent Variable:	ESG Disagreement		
	Coef.	t-stat.	
<i>Time</i>	0.121	(5.98)	***
Intercept	11.283	(65.12)	***
Firm F.E.	Yes		
Adj. R ²	0.486		
N	30,700		

Panel B: ESG disclosure over time

Dependent Variable:	ESG Disclosure		
	Coef.	t-stat.	
<i>Time</i>	1.388	(51.89)	***
Intercept	16.609	(72.45)	***
Firm F.E.	Yes		
Adj. R ²	0.870		
N	30,700		

Panel C: Average ESG performance over time

Dependent Variable:	ESG Avg		
	Coef.	t-stat.	
<i>Time</i>	0.764	(29.85)	***
Intercept	46.256	(210.9)	***
Firm F.E.	Yes		
Adj. R ²	0.872		
N	30,700		

This table reports the trends over time of ESG rating disagreement in Panel A, ESG disclosure in Panel B, and Average ESG performance in Panel C. Time is the difference between the fiscal-year for a particular firm-year observation and 2004, which is the first year in the sample. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 4
Is better ESG Disclosure or Performance associated with greater ESG Rating Disagreement?

Dependent Variable: ESG Disagreement

	OLS			OLS w/ Firm F.E.		
	Coef.	t-stat.		Coef.	t-stat.	
<i>ESG Disclosure</i>	0.127	(16.27)	***	0.094	(10.23)	***
<i>ESG Avg</i>	-0.158	(-19.32)	***	-0.206	(-18.09)	***
<i>Firm Size</i>	0.928	(12.57)	***	-0.012	(-0.06)	
<i>ROA</i>	1.723	(2.23)	**	-0.500	(-0.64)	
<i>BTM</i>	-0.438	(-2.87)	***	-0.290	(-1.97)	**
<i>Leverage</i>	-0.142	(-0.35)		1.710	(2.82)	***
<i>Analyst Following</i>	-0.411	(-3.34)	***	0.072	(0.53)	
<i>Inst. Ownership</i>	-0.010	(-2.60)	***	-0.012	(-1.86)	*
Intercept	14.835	(3.66)	***	13.806	(4.15)	***
ESG Rater F.E.	Yes			Yes		
Year F.E.	Yes			Yes		
Industry F.E.	Yes			No		
Country F.E.	Yes			No		
Firm F.E.	No			Yes		
Adj. R ²	0.118			0.516		
N	30,700			30,700		

This table reports the results of tests based on Equation 1 examining the effects of ESG disclosure on ESG disagreement. There are two specifications: (1) ordinary least squares (OLS) without firm fixed effects, and (2) OLS with firm fixed effects. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 5
Is better Pillar-specific Disclosure associated with greater Rating Disagreement?

Dependent Variable:	<i>ESG_Disagreement</i>				<i>E_Disagreement</i>				<i>S_Disagreement</i>				<i>G_Disagreement</i>			
	OLS		OLS w/ Firm F.E.		OLS		OLS w/ Firm F.E.		OLS		OLS w/ Firm F.E.		OLS		OLS w/ Firm F.E.	
<i>E_Disclosure</i>	0.075 *** (10.14)	0.047 *** (5.95)			0.008 (1.14)	0.026 *** (2.90)										
<i>S_Disclosure</i>	0.018 *** (2.83)	0.036 *** (5.20)						0.08 *** (13.01)	0.058 *** (7.22)							
<i>G_Disclosure</i>	0.039 *** (3.77)	-0.003 (-0.34)										0.036 *** (3.25)	-0.013 (-1.06)			
Intercept	14.729 *** (3.63)	15.003 *** (4.49)			11.258 ** (1.96)	13.534 *** (3.14)			17.414 *** (3.66)	19.033 *** (5.10)			28.029 *** (4.17)	23.929 *** (5.33)		
Controls	Yes	Yes			Yes	Yes			Yes	Yes			Yes	Yes		
ESG Rater F.E.	Yes	Yes			Yes	Yes			Yes	Yes			Yes	Yes		
Year F.E.	Yes	Yes			Yes	Yes			Yes	Yes			Yes	Yes		
Industry F.E.	Yes	No			Yes	No			Yes	No			Yes	No		
Country F.E.	Yes	No			Yes	No			Yes	No			Yes	No		
Firm F.E.	No	Yes			No	Yes			No	Yes			No	Yes		
Adj. R ²	0.119	0.517			0.066	0.432			0.083	0.435			0.283	0.504		
N	30,700	30,700			30,700	30,700			30,700	30,700			30,700	30,700		
F-test (p-value)																
<i>E_Disc=S_Disc</i>	0.000 ***	0.382			-	-			-	-			-	-		
<i>E_Disc=G_Disc</i>	0.010 ***	0.000 ***			-	-			-	-			-	-		
<i>S_Disc=G_Disc</i>	0.108	0.002 ***			-	-			-	-			-	-		

This table reports the effects of disclosure on disagreement for each respective ESG pillar: Environment (E), Social (S), or Governance (G). For each pillar, there are two specifications: (1) ordinary least squares (OLS) without firm fixed effects, and (2) OLS with firm fixed effects. Controls are the same as before, except when an individual pillar is examined (e.g., *E_Disagreement*), then *ESG_Avg* is replaced with the average rating for the pillar being examined (e.g., *E_Avg*). All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 6
Mandatory ESG Disclosure

Dependent Variable:	Validation		Diff-in-Diff		Dynamic Validation		Dynamic Diff-in-Diff					
	ESG_Disclosure		ESG_Disagreement		ESG_Disclosure		ESG_Disagreement					
	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.				
<i>Mandatory_Disclosure</i>	1.435	(5.24)	***	0.448	(2.20)	**						
<i>Mandatory_Disclosure t = -2</i>						0.233	(0.85)	-0.206	(-1.05)			
<i>Mandatory_Disclosure t = -1</i>						0.323	(1.07)	-0.077	(-0.33)			
<i>Mandatory_Disclosure t = 0</i>						1.768	(4.89)	***	-0.366	(-1.30)		
<i>Mandatory_Disclosure t ≥ +1</i>						1.387	(3.48)	***	0.826	(2.72)	**	
<i>ESG_Avg</i>	0.274	(21.45)	***	-0.181	(-15.82)	***	0.273	(21.47)	***	-0.181	(-15.82)	***
<i>Firm Size</i>	0.759	(3.34)	***	0.054	(0.28)		0.752	(3.30)	***	0.070	(0.36)	
<i>ROA</i>	0.201	(0.24)		-0.463	(-0.59)		0.198	(0.23)		-0.430	(-0.54)	
<i>BTM</i>	0.244	(1.29)		-0.272	(-1.84)	*	0.240	(1.27)		-0.274	(-1.85)	*
<i>Leverage</i>	0.089	(0.11)		1.753	(2.84)	***	0.088	(0.11)		1.787	(2.90)	***
<i>Analyst Following</i>	0.725	(4.34)	***	0.145	(1.05)		0.722	(4.31)	***	0.154	(1.13)	
<i>Inst. Ownership</i>	-0.004	(-0.52)		-0.012	(-1.89)	*	-0.004	(-0.50)		-0.012	(-1.91)	*
Intercept	-17.676	(-3.57)	***	12.127	(3.56)	***	-17.601	(-3.56)	***	11.913	(3.50)	***
ESG Rater F.E.	Yes			Yes			Yes			Yes		
Year F.E.	Yes			Yes			Yes			Yes		
Firm F.E.	Yes			Yes			Yes			Yes		
Adj. R ²	0.885			0.511			0.885			0.512		
N	30,700			30,700			30,700			30,700		

This table reports the results using mandatory ESG disclosure requirements, which went into effect in different countries and at different times. The validation columns confirm that mandatory disclosure requirements improve firms' ESG disclosures. The Diff-in-Diff columns show the difference-in-differences analysis. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 7
Multivariate analyses of ESG Metric Disagreement

Panel A: Firm-level ESG Disclosure

Dependent variable: ESG Metric Disagreement

	Baseline		High ESG Disclosure	
<i>Input/Outcome</i>	2.869*** (33.11)	2.726*** (31.47)	2.877*** (22.87)	2.728*** (21.88)
<i>Both Outcomes</i>	0.585*** (5.00)	0.451*** (3.87)	-0.023 (-0.13)	-0.128 (-0.73)
<i>High ESG_Disclosure</i>			1.553*** (13.21)	0.616*** (5.76)
<i>High ESG_Disclosure × Input/Outcome</i>			-0.160 (-0.99)	0.005 (0.03)
<i>High ESG_Disclosure × Both Outcomes</i>			0.748*** (3.41)	0.855*** (3.91)
Intercept	14.651*** (39.64)	10.320*** (10.77)	15.868*** (41.89)	10.329*** (10.80)
Controls	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Industry F.E.	Yes	No	Yes	No
Country F.E.	Yes	No	Yes	No
Firm F.E.	No	Yes	No	Yes
Adj. R ²	0.017	0.056	0.020	0.056
N	927,352	927,352	927,352	927,352

TABLE 7 (continued)

Panel B: Industry-level ESG Disclosure

Dependent variable: ESG_Metric_Disagreement

	Industry ESG Disclosure		Industry & Firm ESG Disclosure	
<i>Input/Outcome</i>	2.952*** (27.00)	2.757*** (25.40)	2.936*** (22.24)	2.744*** (20.95)
<i>Both Outcomes</i>	-0.495*** (-3.03)	-0.660*** (-4.11)	-0.800*** (-4.04)	-0.919*** (-4.67)
<i>High ESG_Disclosure</i>			1.588*** (13.47)	0.651*** (6.07)
<i>High ESG_Disclosure</i> × <i>Input/Outcome</i>			-0.113 (-0.67)	0.023 (0.14)
<i>High ESG_Disclosure</i> × <i>Both Outcomes</i>			0.363 (1.63)	0.458** (2.07)
<i>Industry High ESG_Disclosure</i>	-0.075 (-0.95)	-0.066 (-0.88)	-0.116 (-1.48)	-0.070 (-0.92)
<i>Industry High ESG_Disclosure</i> × <i>Input/Outcome</i>	-0.151 (-0.99)	-0.050 (-0.33)	-0.173 (-1.09)	-0.053 (-0.34)
<i>Industry High ESG_Disclosure</i> × <i>Both Outcomes</i>	1.831*** (8.56)	1.879*** (8.83)	1.753*** (8.04)	1.787*** (8.28)
Intercept	14.705*** (39.37)	10.215*** (10.63)	15.925*** (41.66)	10.252*** (10.69)
Controls	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Industry F.E.	Yes	No	Yes	No
Country F.E.	Yes	No	Yes	No
Firm F.E.	No	Yes	No	Yes
Adj. R ²	0.018	0.056	0.020	0.056
N	927,352	927,352	927,352	927,352

This table shows the regression results of disagreement for each matched-metric from Sustainalytics and ASSET4. The variables of interest are *Input/Outcome* (e.g., where the Sustainalytics metric for the topic is an input and the corresponding ASSET4 metric is an outcome, or vice versa), *Both Outcomes* (i.e., both Sustainalytics and ASSET4 use an outcome metric), and the interaction terms between these variables and ESG disclosure (*High ESG_Disclosure*). The comparison group in the model is *Both Inputs* (i.e., both Sustainalytics and ASSET4 use an input metric). Panel A illustrates the role of firms' ESG disclosure and Panel B adds industry-level ESG disclosure. Controls consists of the same variables used in the prior tests, except now *ESG_Avg* is calculated at the metric level, rather than at the firm-level. Similarly, *ESG_Metric_Disagreement* is calculated at the metric level. See Appendix A for variable definitions. ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

TABLE 8
Consequences of ESG Rating Disagreement using 3-Day Windows

Panel A: Market Consequences

Dependent Variable:	<i>Absolute CAR</i>		<i>Return Volatility</i>		<i>Bid-Ask Spread</i>	
<i>ESG_Disagreement</i>	0.011*** (5.24)	0.008*** (3.42)	0.004*** (4.68)	0.003** (2.57)	0.001 (1.01)	0.000 (0.05)
<i>ESG_Avg</i>	-0.012*** (-7.56)	-0.007*** (-2.63)	-0.005*** (-6.82)	-0.002** (-2.21)	0.000 (0.34)	0.001*** (2.91)
<i>Firm Size</i>	-0.341*** (-26.27)	-0.494*** (-8.07)	-0.141*** (-25.02)	-0.210*** (-7.45)	-0.063*** (-14.19)	-0.068*** (-6.41)
<i>ROA</i>	-4.025*** (-15.11)	-1.691*** (-5.21)	-1.577*** (-13.88)	-0.630*** (-4.34)	-0.306*** (-4.98)	-0.096* (-1.76)
<i>BTM</i>	0.589*** (13.81)	0.562*** (8.33)	0.232*** (13.44)	0.195*** (7.32)	0.093*** (8.52)	0.077*** (5.45)
<i>Leverage</i>	0.774*** (8.96)	1.532*** (7.10)	0.315*** (8.35)	0.614*** (6.10)	0.122*** (5.70)	0.233*** (5.76)
<i>Analyst Following</i>	0.106*** (3.94)	-0.062 (-1.32)	0.040*** (3.30)	-0.026 (-1.26)	-0.058*** (-6.81)	-0.073*** (-7.66)
<i>Inst. Ownership</i>	-0.002*** (-3.84)	-0.004*** (-6.18)	-0.001*** (-2.92)	-0.001*** (-4.09)	-0.001*** (-7.02)	-0.001*** (-7.73)
Intercept	5.665*** (45.66)	6.759*** (11.98)	2.362*** (43.93)	2.876*** (11.23)	0.915*** (26.60)	0.864*** (8.94)
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	No	Yes	No	Yes	No
Country F.E.	Yes	No	Yes	No	Yes	No
Firm F.E.	No	Yes	No	Yes	No	Yes
Adj. R ²	0.135	0.213	0.109	0.179	0.412	0.612
N	74,032	74,032	74,032	74,032	74,032	74,032

Panel B: Market Consequences over Time

Dependent Variable:	<i>Absolute CAR</i>		<i>Return Volatility</i>		<i>Bid-Ask Spread</i>	
<i>ESG_Disagreement</i>	0.004 (0.92)	0.000 (0.01)	-0.000 (-0.24)	-0.002 (-1.13)	0.000 (1.63)	-0.000 (-0.55)
<i>ESG_Disagreement</i> × <i>Time</i>	0.001 (1.49)	0.002* (1.90)	0.001** (2.42)	0.001*** (2.71)	-0.000 (-1.31)	0.000 (0.70)
Intercept	5.684*** (45.38)	6.742*** (12.00)	2.375*** (43.59)	2.865*** (11.25)	0.912*** (26.45)	0.863*** (8.93)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	No	Yes	No	Yes	No
Country F.E.	Yes	No	Yes	No	Yes	No
Firm F.E.	No	Yes	No	Yes	No	Yes
Adj. R ²	0.135	0.213	0.109	0.179	0.412	0.612
N	74,032	74,032	74,032	74,032	74,032	74,032

This table reports the results on the market consequences of ESG rating disagreement using tests during the 3-day window (-1, 0, +1) around the ESG rating release date. Panel A reports the main consequences tests, and Panel B reports the impact of ESG rating disagreement over time. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 9
Consequences of ESG Rating Disagreement on Firm-level Financing Outcomes

Panel A: Financing Outcomes

Dependent Variable:	<i>Equity Issuance</i>		<i>Debt Issuance</i>		<i>Cash Holdings</i>	
<i>ESG_Disagreement</i>	-0.0002 (-0.3465)	-0.0011** (-2.1515)	-0.0022*** (-4.4612)	-0.0000 (-0.0186)	0.0007*** (3.8483)	0.0002* (1.8137)
<i>ESG_Avg</i>	0.0006 (1.4189)	-0.0001 (-0.1166)	-0.0004 (-1.0990)	0.0003 (0.6435)	-0.0001 (-0.3824)	-0.0001 (-1.3991)
<i>Firm Size</i>	0.0134*** (2.7317)	0.0115 (1.1401)	0.0778*** (19.1637)	0.0450*** (4.1696)	-0.0206*** (-11.9057)	-0.0270*** (-9.5946)
<i>ROA</i>	-0.2020*** (-3.8529)	0.1445*** (2.7812)	-0.0969* (-1.8560)	0.0854 (1.5593)	-0.0927*** (-3.6930)	0.0006 (0.0512)
<i>BTM</i>	-0.0884*** (-8.4021)	-0.0745*** (-6.6218)	-0.0126 (-1.3729)	-0.0581*** (-6.6726)	-0.0414*** (-12.0585)	-0.0049** (-2.5095)
<i>Leverage</i>	-0.0531* (-1.9212)	0.0272 (0.7102)	0.3382*** (13.2584)	-0.1142*** (-2.9389)	-0.1122*** (-10.0653)	-0.0298*** (-2.9573)
<i>Analyst Following</i>	0.0163** (2.0216)	0.0043 (0.4962)	-0.0291*** (-4.4479)	0.0225*** (2.7661)	0.0244*** (8.8754)	-0.0009 (-0.5056)
<i>Inst. Ownership</i>	0.0018*** (6.5315)	0.0000 (0.0700)	0.0005** (2.2628)	-0.0003 (-0.8191)	-0.0006*** (-6.0933)	-0.0000 (-0.2141)
Intercept	0.4067*** (10.4977)	0.5438*** (5.4594)	-0.0818** (-2.3413)	0.2994*** (2.8977)	0.3760*** (26.2289)	0.4045*** (15.8940)
ESG Rater F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	No	Yes	No	Yes	No
Country F.E.	Yes	No	Yes	No	Yes	No
Firm F.E.	No	Yes	No	Yes	No	Yes
Adj. R ²	0.227	0.565	0.154	0.450	0.302	0.835
N	30,700	30,700	30,700	30,700	27,084	27,084

Panel B: Financing Outcomes over Time

Dependent Variable:	<i>Equity Issuance</i>		<i>Debt Issuance</i>		<i>Cash Holdings</i>	
<i>ESG_Disagreement</i>	-0.0017 (-1.3578)	-0.0024** (-2.0097)	0.0010 (0.9111)	0.0016 (1.4683)	0.0001 (0.1846)	-0.0000 (-0.2043)
<i>ESG_Disagreement</i> <i>× Time</i>	0.0002 (1.2709)	0.0002 (1.1796)	-0.0004*** (-3.0591)	-0.0002 (-1.6414)	0.0001* (1.7687)	0.0000 (1.0437)
Intercept	0.4086*** (10.5578)	0.5404*** (5.4216)	-0.0858** (-2.4559)	0.3035*** (2.9404)	0.3768*** (26.2745)	0.4040*** (15.8550)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
ESG Rater F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	No	Yes	No	Yes	No
Country F.E.	Yes	No	Yes	No	Yes	No
Firm F.E.	No	Yes	No	Yes	No	Yes
Adj. R ²	0.227	0.565	0.154	0.450	0.302	0.835
N	30,700	30,700	30,700	30,700	27,084	27,084

This table reports the results of ESG rating disagreement on firm-level financing outcomes. Panel A reports the main financing consequences tests, and Panel B reports the impact over time. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings

Online Appendix

This document contains additional tests that are untabulated in the paper. These analyses include: (1) performing univariate tests on the relationship between ESG disclosure and ESG disagreement, and (2) examining the role of weighting schemes in ESG disagreement.

OA1. Univariate tests

We perform two univariate tests on the relationship between ESG disclosure and ESG disagreement. First, we calculate Pearson correlations between our measures ESG disclosure and ESG disagreement. As reported in Table OA1, we find that ESG disclosure is positively correlated with ESG disagreement, consistent with our hypothesis.

Next, we perform portfolio analyses, where we allocate firms into different portfolios. Specifically, we create 3x3 portfolios, based on a firm's level of ESG disclosure and average rating of ESG performance. To create these portfolios, we rank *ESG_Disclosure* into terciles (high, medium, and low), and also rank *ESG_Avg* into terciles (high, medium, and low). This yields nine portfolios. For each portfolio, we tabulate the mean and median ESG rating disagreement in Table OA2 Panels A and B, respectively. Again, consistent with our main hypothesis, across all the portfolios we see that disagreement increases monotonically as disclosure increases, consistent with our hypothesis.

A few additional insights are also revealed by this table. Specifically, we also find that disagreement is highest for the firms rated as the worst ESG performers. We also observe a U-shaped relationship between ESG disagreement and average ESG ratings, as the lowest level of

disagreement is found for the firms in the middle tercile.¹

OA2. Pillar Weights

In the paper we have focused our analyses on the importance of disclosure as an explanatory variable. Another thing that could play a role in rating disagreement is disagreement on the weights raters place on the three ESG pillars. To test this, we created a weight disagreement measure for each firm-year by calculating the standard deviation of weights that raters used for each ESG pillar (e.g., environmental weight disagreement), and then took the average of the weight disagreement scores for the E, S, and G pillars (*ESG_Weight_Disagree*). As expected, the univariate correlation between this measure and ESG rating disagreement is positive and statistically significant ($p < 0.05$; untabulated). However, when we include this variable in our multivariate analyses, as reported in Table OA3, *ESG_Weight_Disagree* is not statistically significant.² Thus, its influence on disagreement appears to be subsumed by the other variables in the model. Because most raters place different weights on the pillars by industry (e.g., MSCI 2018), with some deviations to account for idiosyncratic firm cases, it is not too surprising that our primary results are unaffected by the inclusion of this variable, as the model includes industry fixed effects. However, this analysis does not necessarily imply that weight disagreement is unimportant, as the importance of an industry-level effect could be subsumed in our models where the unit of analysis is the firm-year.

To shed more light on this issue, we move to an industry-year level analysis to understand whether weight disagreement explains rating disagreement across industries. We calculate all

¹ This U-shaped relationship holds in a multivariate setting as well (untabulated). Specifically, when we create an indicator variable *High ESG_Avg* (*Low ESG_Avg*) equal to one if a firm-year is in the highest (lowest) average ESG performance tercile, and then re-run Equation 1 where we replace *ESG_Avg* with these two indicator variables, we observe that coefficients on both *High ESG_Avg* and *Low ESG_Avg* are positive and statistically significant.

² The results are similar if we instead just include the three pillar weight disagreement scores as separate variables in the model (instead of aggregating them into a single variable).

variables at the industry-year level by averaging firm-year observations within each industry-year. The univariate correlation between rating disagreement and weight disagreement for the industry-year observations is positive and statistically significant ($p < 0.01$; untabulated). Industries with highest weight disagreement are personal services, utilities, and chemicals. When we perform a multivariate analysis, which is also reported in Table OA3, we see that the industry average weight disagreement is positively associated with the industry average ESG rating disagreement ($p < 0.05$). Thus, at the industry-year level, disagreement about what weights to assign each ESG pillar does help explain some of the disagreement in ESG ratings. Overall, the results suggest that weight disagreement does not affect our primary inferences about firm-level variation in disagreement, but could play a role in understanding industry-level variation in disagreement.

TABLE OA1
Pearson Correlations

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1. <i>ESG_Disagreement</i>																	
2. <i>E_Disagreement</i>	0.36																
3. <i>S_Disagreement</i>	0.53	0.15															
4. <i>G_Disagreement</i>	0.25	0.02	0.08														
5. <i>ESG_Avg</i>	-0.05	-0.05	-0.04	-0.12													
6. <i>E_Avg</i>	-0.01	0.00	0.02	0.06	0.86												
7. <i>S_Avg</i>	-0.05	-0.03	-0.07	-0.02	0.90	0.75											
8. <i>G_Avg</i>	-0.04	-0.04	-0.07	-0.39	0.60	0.28	0.36										
9. <i>ESG_Disclosure</i>	0.12	0.00	0.11	0.10	0.70	0.71	0.68	0.23									
10. <i>E_Disclosure</i>	0.12	0.00	0.12	0.15	0.65	0.70	0.64	0.14	0.97								
11. <i>S_Disclosure</i>	0.09	-0.01	0.09	0.06	0.65	0.62	0.66	0.23	0.90	0.79							
12. <i>G_Disclosure</i>	0.09	0.03	0.06	-0.14	0.50	0.38	0.39	0.47	0.60	0.46	0.49						
13. <i>Firm Size</i>	0.11	0.01	0.07	0.07	0.33	0.35	0.35	0.03	0.38	0.38	0.30	0.24					
14. <i>ROA</i>	0.02	-0.03	-0.01	0.01	0.03	0.03	0.04	0.04	0.01	0.00	0.02	0.02	-0.16				
15. <i>BTM</i>	0.00	-0.02	0.02	0.09	-0.02	0.01	0.03	-0.15	0.08	0.10	0.07	-0.04	0.28	-0.35			
16. <i>Leverage</i>	0.02	0.02	0.00	-0.06	0.12	0.08	0.13	0.07	0.08	0.06	0.07	0.09	0.48	-0.25	-0.02		
17. <i>Analyst Following</i>	0.02	0.01	-0.02	-0.03	0.38	0.35	0.36	0.21	0.32	0.30	0.29	0.26	0.32	0.13	-0.13	0.06	
18. <i>Inst. Ownership</i>	-0.05	0.04	-0.08	-0.26	0.04	-0.12	-0.13	0.43	-0.20	-0.24	-0.23	0.22	-0.15	0.04	-0.25	0.02	0.09

This table presents correlations for the variables in our tests (N=30,700). Bolded correlations are significant at the 0.05 level or lower. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions.

TABLE OA2
Portfolio Analyses

Panel A: Mean ESG Disagreement

		<i>ESG Avg</i>			
		Low	Medium	High	Diff
<i>ESG_Disclosure</i>	Low	12.88	8.92	11.43	-1.45
	Medium	14.23	9.83	12.72	-1.51
	High	17.51	12.03	13.69	-3.82
	Diff	4.63	3.11	2.26	

Panel B: Median ESG Disagreement

		<i>ESG Avg</i>			
		Low	Medium	High	Diff
<i>ESG_Disclosure</i>	Low	12.31	7.87	10.53	-1.78
	Medium	13.32	8.76	12.73	-0.59
	High	17.88	10.91	13.55	-4.33
	Diff	5.57	3.04	3.02	

This table presents univariate results from portfolio analyses of ESG rating disagreement. For each year we allocate each firms into terciles of ESG disclosure and terciles of the average ESG score across the rating agencies. Panel A (B) shows the mean (median) ESG disagreement across ratings for each of the nine portfolios of companies. Difference is calculated as High minus Low.

TABLE OA3
Multivariate association between ESG Rating Disagreement and ESG Weight Disagreement

Dependent Variable: ESG Disagreement

	Firm-Year Test			Industry-Year Test		
	Coef.	t-stat.		Coef.	t-stat.	
<i>ESG Weight Disagree</i>	2.125	(1.30)		17.948	(2.10)	**
<i>ESG Disclosure</i>	0.127	(16.28)	***	0.142	(1.93)	*
<i>ESG Avg</i>	-0.158	(-19.32)	***	-0.167	(-2.51)	**
<i>Firm Size</i>	0.933	(12.62)	***	0.090	(0.23)	
<i>ROA</i>	1.713	(2.22)	**	13.637	(1.05)	
<i>BTM</i>	-0.440	(-2.88)	***	0.382	(0.31)	
<i>Leverage</i>	-0.154	(-0.38)		-0.669	(-0.22)	
<i>Analyst Following</i>	-0.413	(-3.36)	***	0.599	(0.90)	
<i>Inst. Ownership</i>	-0.010	(-2.58)	***	0.003	(0.13)	
Intercept	11.847	(7.87)	***	12.999	(3.91)	***
ESG Rater F.E.	Yes			No		
Year F.E.	Yes			Yes		
Industry F.E.	Yes			No		
Country F.E.	Yes			No		
Firm F.E.	No			No		
Adj. R ²	0.118			0.131		
N	30,700			576		
Unit of observation	Firm-Year			Industry-Year		

This table reports the results on the relationship between ESG rating disagreement and ESG weight disagreement using tests at the firm-level and industry-level. *ESG Weight Disagree* is calculated by first taking the standard deviation of weights that raters used for each ESG pillar (e.g., environmental weight disagreement), and then taking the average of the weight disagreement scores for the E, S, and G pillars. See Appendix A for all other variable definitions. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.